THE SCALING OF TEACHERS' MARKS AND ESTIMATES

THE SCALING OF TEACHERS' MARKS AND ESTIMATES

By

DOUGLAS M. McINTOSH
M.A., B.Sc., B.Ed., Ph.D.
Director of Education, Fife.

DAVID A. WALKER
M.A., B.Ed., Ph.D.
Depute Director of Education, Fife.

DONALD MACKAY
M.B.E., M.C., M.A.
Headmaster, Viewforth Secondary School, Kirkcaldy.

OLIVER AND BOYD
EDINBURGH: TWEEDDALE COURT
LONDON: 98 GREAT RUSSELL STREET, W.C.

OLIVER AND BOYD LTD.
EDINBURGH AND LONDON

Sole Agents outside the British Isles
(Except in Canada and the U.S.A.)
MACMILLAN & CO. LIMITED
LONDON BOMBAY CALCUTTA MADRAS MELBOURNE

Sole Agents for Canada
CLARK, IRWIN & CO. LIMITED
TORONTO

1949

PRINTED IN GREAT BRITAIN BY
A. WALKER & SON, LTD., GALASHIELS
FOR OLIVER & BOYD LTD., EDINBURGH

FOREWORD

THE publication of this book reflects a belief, which I share with Dr McIntosh, that the scaling of teachers' marks will have an established and permanent place not merely in educational research but in many departments of everyday educational practice. We used the technique on a large scale in our Dundee investigation on Selection for Secondary Education, and our experience left us both with a conviction of its practical possibilities—a conviction that has been greatly strengthened by the results of later researches by Dr McIntosh, Dr Walker and Mr Mackay.

Estimates of pupils' attainments in school subjects are needed for a multitude of practical educational purposes; and it is part of my educational faith that the assessments of the teachers provide the best guide that is, or ever will be, available as to the *order of merit* of their pupils in these attainments. But, unfortunately, the standard and scatter of marks vary from teacher to teacher and from school to school; and while teachers' assessments may be improved in this respect, it is extremely doubtful whether they could ever be steadied to such an extent that they could be used with assurance as a basis for the serious decisions involved in selection for secondary education or the award of Leaving Certificates. If these premises are sound, there is no escape from the conclusion that teachers' marks should be used for such purposes, but only if they have been scaled in such a way that the estimates of different teachers are comparable.

Scaling has already been employed in practice to a considerable extent, particularly in selection for secondary education, but one suspects that it would have been more widely used but for two relatively groundless fears. The first is that this is a tool that can be handled only by the expert mathematical statistician: the second is that a prohibitive amount of laborious calculation is involved. The present volume should do much to dispel these fears. It gives a simple explanation

of the basis of scaling, and clear and definite instructions as to the best ways of making the calculations. Now that such a work is available there should be little difficulty in the use of the technique by the staff of an Education Office, the headmaster or the class teacher.

The book is, however, much more than a statistical primer giving instructions for the use of methods that are already well known: it is an important contribution to educational research. In the hands of Dr McIntosh and his colleagues, new and improved scaling procedures have been evolved—procedures that have passed the acid test of extensive use in actual practice. New fields for the application of scaling methods are opened up; and headmasters and class teachers who study the authors' suggestions will find their reward in increased interest in their work and fuller satisfaction that they are carrying out their important task of adjusting the education they provide to the varying talents and needs of their pupils.

WILLIAM MCCLELLAND

PREFACE

THE ideas developed in this book have their origin in McClelland's "Selection for Secondary Education." In that book a clear case was presented against the use of teachers' estimates for the purpose of selection unless they were made comparable from school to school.

These ideas were further developed in the Report of the Scottish Advisory Council on Secondary Education. Here the scaling procedure was suggested for use in an examination scheme for the award of the Leaving Certificate and, again, on the discussion of marks and terminal reports, it was suggested that school examination marks should be made comparable by some method not aiming at " statistical impeccability, but rather something reasonably valid yet easily workable by the busy teacher in the school."

We have attempted to attain the Advisory Council's objective. No deep knowledge of statistics is required for the understanding of our methods. Chapter III on "A Little Statistics " will, we hope, give all that is necessary for the understanding of the techniques we have devised: it is an attempt at statistics without tears.

We should like to emphasise also that our methods are not merely theoretical but have been adopted in practice both in education offices and in a large junior secondary school. The needs of the administrator and " the busy teacher in school " have been in the forefront of our minds all the time.

We wish to acknowledge our indebtedness to Mr McClelland for writing the foreword, for the keen interest he showed in our work and for the many helpful suggestions which he made. Mr W. F. Lindsay, Assistant Director of Education for Fife, read the manuscript in its various stages and offered many constructive criticisms. We also received much help regarding the layout of the book from Mr S. Stewart, Educational Editor of Messrs Oliver & Boyd. For illustrative materials in Chapters VII and VIII we are indebted to Midlothian Education Committee.

CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

EXAMINATIONS

EXAMINATIONS have been for long one of the most controversial topics in education. There are those who hold strongly that examination results are the only true measure of the success or failure of education and they do not hesitate to cut out " frills " such as physical training and music as soon as examination day is in sight. On the other hand there are those who claim that most of the faults of present day education can be traced to the examination system; education is a thing of the spirit and to examine subjects such as literature is " as impossible as to imprison sunbeams." No doubt the truth lies somewhere between these two extreme points of view; examinations are certainly not the " Alpha and Omega of education," yet on the other hand " the life without examination is a life that can hardly be lived."

There are few who would argue that class tests or examination are harmful if they are not too frequent. A teacher must have some check whether a pupil has been profiting by instruction; for example, if a teacher wishes to know whether his pupils can add or subtract, he must give them several addition and subtraction sums to work, and from the results he can tell which pupils have a thorough grasp of these two operations. Each examination or test should be set with a definite aim in view and should be constructed to achieve this aim. For example, a test may be designed to measure a child's level of attainment or may be set with the intention of finding out where his weakness in a particular subject lies. Again, an examination to select a few clever pupils will be quite different from an examination, the aim of which is to pick out a few less able pupils.

Most of the criticism against examinations refers to the external examination. Not only has it " cabin'd, cribb'd, confined " the teacher but its results have often been beyond

the teacher's comprehension: pupils who should have passed with ease fail and some who should have failed pass.

A great volume of research has been carried out on examinations. It has been shown that two examiners will give widely different marks to the same paper and that the same examiner will give quite a surprising variation of marks when he assesses the same pupils at different intervals.

The following quotations illustrate the type of result obtained in these researches:—

" The facts of a subjective scale are well illustrated in the following anecdote concerning the grading of history papers by a group of college professors of history in the summer of 1920. One of the five or six expert readers assigned to a certain group of history papers, after scoring a few, wrote out for his own convenience what he considered a model paper for the given set of questions. By some mischance this model fell into the hands of another reader who graded it in a perfectly bona fide fashion. The mark he assigned to it was below passing, and, in accordance with the custom, this model was read by a number of other expert readers in order to ensure that it was properly marked. The marks assigned to it by these readers varied from 40 to 90." *

French investigators dealing with essays selected three scripts, Nos. 23, 25 and 34, " each of which at the original ' baccalaureat ' examination had been awarded 36 marks out of 80 (or 45 per cent.) and had been ranked as 24th out of a batch of 50. These three scripts were marked independently by 76 examiners. The marks for script No. 23 varied from 4 to 52, for script No. 25 from 12 to 64, and for script No. 34 from 16 to 65 out of a maximum of 80. The mean marks for the three scripts were as follows:—

Script No. 23—25·9; Script No. 25—40·0; Script No. 34—34·4."

In an English experiment 14 examiners were asked to re-mark 15 history scripts some 12 to 19 months later, having kept no record of their previous marks. The examiners

* Examinations and their Substitutes in the United States: The Carnegie Foundation for the Advancement of Teaching, Bulletin No. 28, 1936, page 64.

awarded not only numerical marks but the verdict of Failure, Pass or Credit. "It was found that in 92 cases out of the 210 the individual examiners gave a different verdict on the second occasion from the verdict awarded on the first." *

These results refer, of course, to the essay type of examination. To avoid such variations the objective type of question was devised where there is only one correct answer and this receives a definite mark. Examples of this type of question are:—

1. Underline the word that means the opposite of the first word.

   Cautious—guarded, adverse, harsh, rash
   Sympathy—antipathy, passion, disrespect, courtesy

2. Change the following sentence to the *past* tense:—

   The tree bears much fruit (Present)
   The tree. . . . . .much fruit (Past)

3. Complete the following:—

$$\frac{9}{13} = \frac{\quad}{117}$$

Objective tests have been devised chiefly for primary school subjects. There is a growing volume of experiment with tests of this type for the secondary school curriculum. In Appendix I we give sample questions from tests devised for pupils at the Leaving Certificate stage.

The great advantage of tests devised in this way is that any two teachers correcting them will be sure to award the same mark to each script: there can be no variation in the standard of marking. Considerable ingenuity has been shown in the construction of such tests and they can be made to test not only attainment but a child's ability to reason.

The disadvantage of these tests is that they cannot easily be made to measure a pupil's ability to marshall ideas and set them down in an orderly sequence. This ability may be a very important part of the subject: in English, for example, composition must always play an important part

* HARTOG & RHODES. *An Examination of Examinations.* International Institute Examinations Enquiry, page 81 and page 15. (Macmillan, 1935).

even though its assessment does present considerable difficulties.

The problems connected with setting and evaluating examinations are not primarily our task. It is necessary, however, to stress the desirability of improving examination techniques for the very good reason that it is not worth while to go through the scaling process with marks that are not themselves good estimates of the pupils' achievements. Scaling does nothing to add to the reliability of marks; its function is to give them meaning and to render them comparable.

CHAPTER II

THE NEED FOR SCALING

In the Advisory Council's Report on Secondary Education* this challenging sentence occurs:—"It is important that the records of any school should be in a form that means something to another school or to an outside body and that the report to the parent should enlighten and not mislead." The implication of this statement is that at present marks of one school do not mean much to another school, and that parents get their reports in a form which leaves them very much 'in the dark.' Coming nearer home it might be asked whether, in a given school, class marks mean something to all members of the staff and whether that something is the same thing.

An examination mark by itself has practically no meaning; the traditional belief that it has an absolute value, so many per cent. of a possible or "perfect" performance, is without foundation. All that a mark of 60 per cent. conveys is that it is one of a set of marks which has a range within the limits zero to one hundred. Whether the mark is to be reckoned as "very good," "good" or "not so good" ("high," "medium" or "low") depends on how 60 is related to the other marks made by the class.

When the teacher has finished marking a set of examination papers he has a clear idea of the value of the mark 60. If he puts the papers in order of merit and counts down to find the mark of the middle pupil he is only making explicit what was implicit in his mind before. If he goes further and finds the mark of the pupil who is one quarter of the way from the top and bottom respectively, he will get a still clearer picture of the value of 60. In the teacher's marks book the marks retain, at least for a time, their full meaning. Later on they may be transferred to the class teacher's marks register, to a head-

* *Secondary Education.* A Report of the Advisory Council on Education in Scotland, Chap. XIII, para. 700, page 147. (H.M. Stationery Office, 1947).

teacher's record card and to a parent's progress card. All that is usually entered on the pupil's permanent record is the entry "English 60." The collateral information that gave 60 its meaning is thrown away and the headteacher is left to wonder if a pupil who has scored 60 in English and 60 in history has done equally well in the two subjects. He has to go on the assumption that the two marks are equal, although only by chance will this be true. It is a serious matter that marks in the files of the headteacher, on which the final assessment of the pupil for educational and vocational guidance or other external purposes is based, should be so lacking in exact specification.

In some secondary schools, in addition to the pupil's mark, the class average or the pupil's place in the order of merit is given against each subject in the report to parents. Even so, it will not be an easy task for the parent to decide how much better or worse 60 in English is than 60 in history. Here is a term report for a pupil from a secondary school:—

Subject	Mark	Class Average
English	61	58
History	70	52
Geography	70	52
Arithmetic	46	48
Algebra	55	56
Geometry	71	63
Science	62	63
French	50	40
Latin	72	68
Art	58	57

What is the parent to make of the fact that in this report the class average mark in French is 40 and in Latin 68? The teacher of Latin knows that he has set a comparatively easy paper and the French teacher recognises that his paper has been too stiff; the headteacher may come to the conclusion (quite unwarrantedly) that the class is better at Latin than at French or that it is a way these two teachers have of marking which he must keep in mind when forming a judgment. The parent, however, taking the figures at their face value, will

come to the conclusion that the French teacher is a poor teacher.

Furthermore, the parent would be misled if he came to the conclusion that his boy was doing equally well in history and geography because, the class average of the two subjects being 52, his boy has 70 marks in each. He would be right only in the odd case when the scatter of marks in the two sets was the same. To add a further column showing that he was ninth in history and fourth in geography, instead of conveying further information, would probably puzzle him. Even the best progress cards at present although they go some way to eliminate the misleading qualities of raw marks, do not fully enlighten.

Marks in the same subject in different terms do not have the same meaning. For example, 60 in the first term may indicate a poorer performance than 60 in the third term. Here is an extract from a pupil's record card:—

Term	1st year			2nd year	
	I.	II.	III.	I.	II.
Mark	60	74	60	79	55
Class Average	59	56	40	65	55

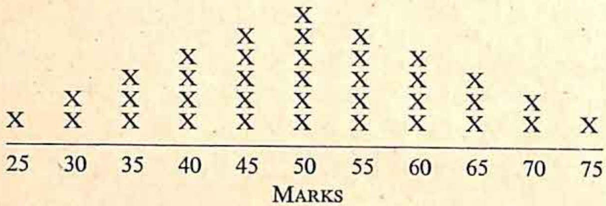
It would appear that the third term 60 is a better performance than the first term 60.

How does the second term 74 compare with the second year first term 79? This question can only be answered if the remainder of the class marks is known or at least if the scatter or spread of the marks is known.

The same problems face the Director of Education who wishes to use Teachers' Estimates in a scheme of selection for secondary courses of instruction: it will be a very lucky chance if a mark of 70 in arithmetic from school X means the same as 70 from school Y.

The points made in the preceding discussion may be illustrated by the following examples. Suppose the English papers of a class of 36 showed the following distributions in their autumn term test (A), each X representing one pupil.

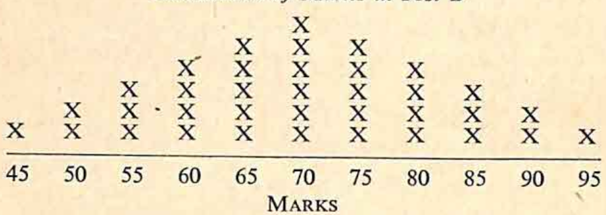
FIG. I
Distribution of Marks in Test A



A pupil with 50 will be right in the middle of his class while a pupil with 70 will have only one of his classmates scoring a higher mark.

Suppose that the spring test (B), considerably easier or marked with less severity, gives the following distribution of marks:—

FIG. II
Distribution of Marks in Test B



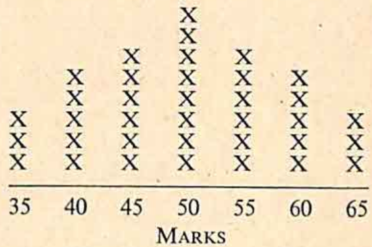
Here a pupil with a mark of 70 will be in the middle of his class. The two 70s have entirely different meanings.

As the scatter of marks is the same in both tests the marks in the spring test may be made directly comparable with those in the autumn test by subtracting 20 from each spring test mark. On the other hand, if we want to combine the marks we do not need to adjust them to the same average. Take an extreme case: if the boy who scores 75 in the autumn (A) test falls to the average in the spring (B) test he gets a total of 145; if the "average" boy in the A test jumps into top place in the B test he also gets 145. The higher average gives no greater "weight" to the spring test—a point that is not generally understood.

Suppose that in a third test (C) the marks are distributed as follows:—

FIG. III

Distribution of Marks in Test C



The average score of 50 in C is the same as in A, but the scatter of marks is much less—the ratio is roughly 30:50. When it comes to combining the three tests, one effect of this shrinkage of scatter in the last term is to penalise the boy who came out top in C by comparison with his classmates who who were first in A or B. Being highly placed in the test scored by the marker of restricted range pays a poor dividend when the annual class list comes out. To illustrate this effect let us suppose a boy was first in A and dropped to the average place in C. Compare him with his opposite number who jumped, we shall suppose, from average place in A to top in C. The former is credited with a total of 125 and the latter with 115. This is obviously unfair; to make things fair the C marks should have been stretched out from the centre both ways so as to increase the range of C to 50.

The general principle is that the greater the scatter of marks the greater the weight of the mark. Differences of scatter, generally overlooked, are just as important as differences of average in interpreting marks. Discrepancies in both should be ironed out if marks are to be compared or combined. When the day comes that a headteacher or Director of Education finds on his desk only marks that have been ironed out or adjusted or scaled to have the same average and scatter, he will then be able to make well-informed judgments on the standing of the pupils.

The extract from a pupil's record card on page 7 shows how widely apart class averages may be in the ordinary run of school marks. Would we find similar differences in scatter?

The marks shown in diagrammatic form in Fig. IV came from a Group B class in a secondary school. The class average marks were controlled to some extent by exhortation; the teachers were asked to aim at an average round about 56. No attempt was made to adjust any realised discrepancies.

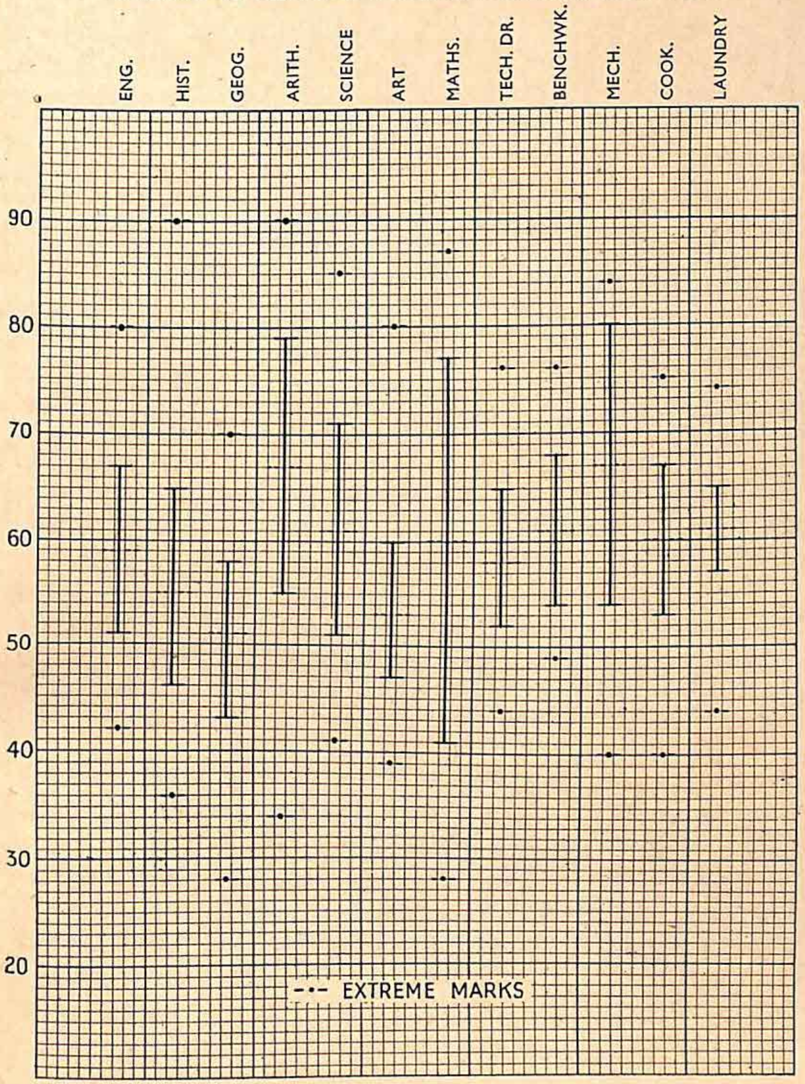
The broken lines show the vagaries of the class average. It jumps up and down as we pass from subject to subject. A measure of scatter may be taken as the difference between the pupil a quarter of the way from the top of the class and the pupil a quarter of the way from the bottom of the class, and this is shown by the length of the vertical lines. This length varies from subject to subject and shows how unequal the scatter of marks has turned out. Clearly, it would pay a boy with his eye on his yearly average mark to score well in mathematics and arithmetic, while the best endeavours of the practical boy depending on his benchwork mark would avail him little.

It can be computed in this series of examinations that 70 in English is really just as high as 83 in arithmetic or 75 in science and no higher than 68 in history, 61 in geography and 62 in art.

Unquestionably, raw marks are in no fit state to be either compared, compounded or interpreted. It is the business of the scaling process to make the necessary transmutations.

FIG. IV

Illustration of Differences in Mean and Scatter of School Marks



CHAPTER III

A LITTLE STATISTICS

An apology—" To those unfamiliar with the recent course of educational psychology the intrusion of mathematical refinements into problems of the classroom may seem to be little short of perverse pedantry, more ridiculous and less excusable than the crotchets of the trigonometrical tailor who fitted Gulliver with a suit of clothes by means of a sextant and a theodolite. Examining children, it may be urged, is among the simplest of the teacher's daily duties; and needs little else but common sense and rule of thumb: by many words, still more by many figures, counsel is darkened . . . .

"While Swift was gibing at the mathematicians of Laputa, Newton was writing his Principia. The satirist may be read the more widely: but the mathematician has more profoundly changed and furthered the course of civilisation." *

This chapter is introduced to give the non-mathematical readers some idea of elementary statistical terms. Lest they take fright we would remind them of Shepherd Dawson's claim that " the calculation of statistics is a matter of mere arithmetic and the understanding of their general significance requires little more than common sense."

Distribution of Marks

After he has marked a set of examination papers a teacher generally arranges them in order, the highest mark on top and the lowest at the bottom. With the papers thus arranged, he can readily give each pupil an ' order of merit ' or a ' rank.' Here are the marks in an English examination for a class of 20 pupils:—

84, 78, 74, 72, 66, 66, 65, 59, 56, 56, 54, 54, 49, 49, 47, 43, 41, 41, 40, 32.

* Sir Cyril Burt. *Mental and Scholastic Tests*, page 142. (Staples Press Ltd., 1947).

The pupil who scores 72 is fourth in order of merit or has the rank 4, the pupil with 59 is eighth or has the rank 8, and so on. Some difficulty is experienced with the mark 66. The teacher generally states that the pupils with this score are fifth equal; the statistician would give each the rank 5·5. If the sixth, seventh and eighth pupils from the top have the same mark, each pupil is given the rank 7.

The results of an external examination dealing with some hundreds of pupils can hardly be treated in this way. In such cases the marks are generally arranged in groups or "class intervals." A simple case is that illustrated by Fig. I, page 8. For ease, we illustrated the case where one pupil scored 25, two pupils 30, three pupils 35, and so on. This is a most unlikely result in any examination. If, however, the one pupil had made a score between 21 and 25, two a score between 26 and 30 and so on, we have a much more likely result. The group of marks 21-25 is known as the *class interval*. The number of cases in a class interval is known as the *frequency*.

The frequency distributions for tests A, B and C are shown below.

TABLE 1

Frequency Distribution of Marks for Examinations A, B and C

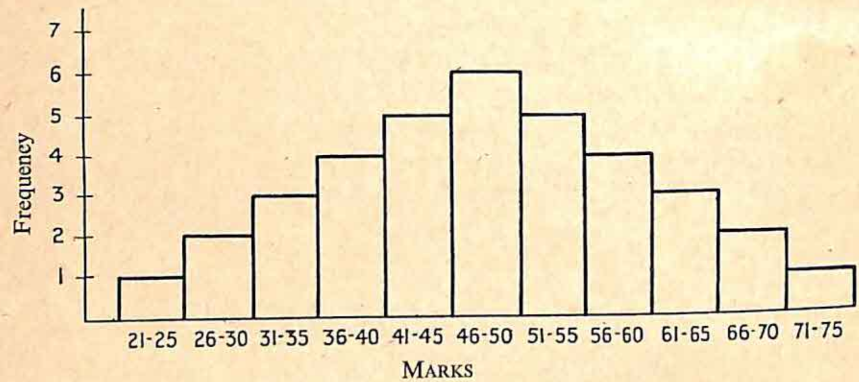
Class Interval	Frequency A	B	C
91-95		1	
86-90		2	
81-85		3	
76-80		4	
71-75	1	5	
66-70	2	6	
61-65	3	5	3
56-60	4	4	5
51-55	5	3	6
46-50	6	2	8
41-45	5	1	6
36-40	4		5
31-35	3		3
26-30	2		
21-25	1		
Total	36	36	36

One of the advantages of this method of recording marks is that it is possible to obtain from the distribution a rough idea of how the test has spaced out the candidates. For this reason the distribution is frequently called the *score scatter*.

This may be further illustrated by the use of graphical representations of score scatters. For example, the distribution of marks shown in column A in Table 1 is shown as follows:—

FIG. V

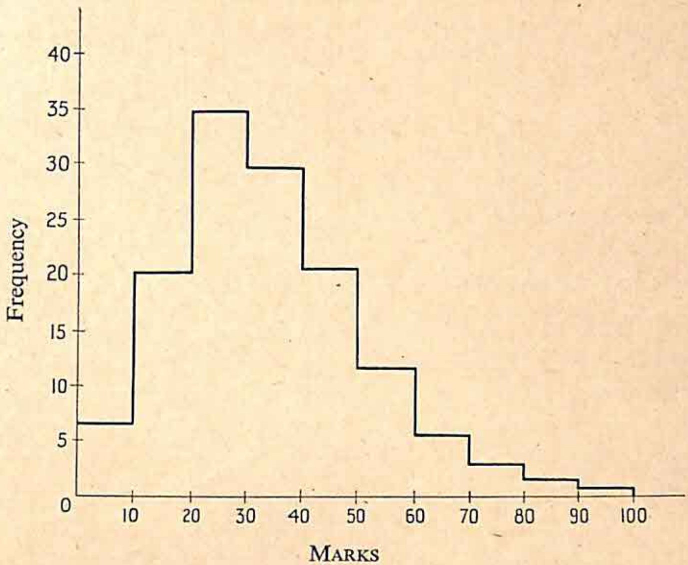
Diagrammatic Representation of Frequency Distribution A



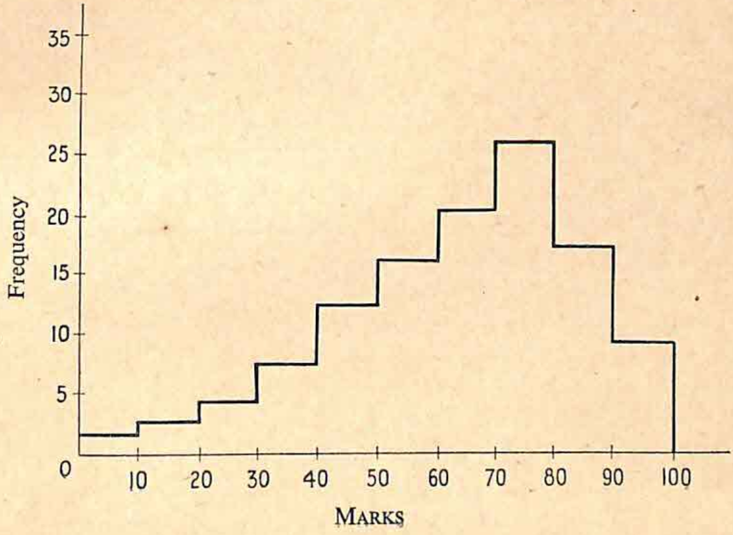
The above diagram is known as a *histogram*.

It is interesting to think out the types of distribution that are most suitable for examinations of different types. For example, an examination designed to select the most able pupils should have a score scatter of the first type shown opposite.

It is then much easier to decide which are the best candidates; the bunching of the pupils at the foot of the scale does not matter.



Conversely an examination for a certificate to be awarded to three-quarters of the candidates should have a score scatter of this type:—



Here the important part of the mark scale is that dividing the lower quarter of the candidates from the upper three quarters.

The devising of examinations to produce these score scatters is a separate subject, on which we shall not enter. We strongly recommend that for each test or examination the teacher sets he should find the distribution of marks.

Measure of Standard—The Average or Mean

The standard of marking is simply expressed by the mean or average mark. For example, the standard of marking in examinations A and C is the same because the mean or average in each case was 50. Examination B was marked on a much easier standard as the class average or mean mark was 70.

Where the number of marks is small the mean or average is easily calculated by adding them and dividing by the number of marks. Where the number of marks is large it is usually better to prepare a score scatter as shown in the previous section and to employ the method of calculation shown in Appendix II.

Measures of Scatter or Spread of Marks

(*a*) *Range*

One simple method of indicating the scatter of marks in a test is to give the difference between the largest and smallest marks. This is known as the *range*. In the three examples above, the ranges are 50, 50 and 35 showing that examination C gave the smallest spread of marks. The range as a measure of scatter is of little use when the marks are not evenly distributed. For example, the range of 50 obtained in test A would be drastically altered if the mark of a single pupil were altered from 75 to 90.

(*b*) *Interquartile Range*

The upper quartile, generally denoted by Q_3, is the mark below which three quarters of the candidates fall and the lower

quartile, denoted by Q_1, is the mark below which one quarter of the candidates fall. The difference between these marks is the interquartile range.

For example, in Table 1, the upper quartile for examination A is the mark below which 27 of the candidates fall. This mark must lie somewhere between 56 and 60 because 30 candidates have scores less than 61 and 26 candidates have scores less than 56. It can be shown to be 57. Similarly the lower quartile is in the range 36-40 and can be shown to be 39. The interquartile range is therefore $57 - 39 = 18$.

(*c*) *Interpercentile Ranges*

The idea of the quartile may be extended to cover not only quarters but also hundredths. The mark below which lies 80 per cent. of all the marks is known as the 80th percentile. In general, a percentile is the mark below which lies a certain percentage of all the marks.

It will be obvious that the 25th percentile is the lower quartile, and the 75th percentile is the upper quartile. The 50th percentile, below which lies half of the candidates, is called the *median*. In this book we shall often use the interpercentile range from the 16th to the 84th percentile as a suitable measure of spread.

We denote the xth percentile by P_x; for example, the 80th percentile is denoted by P_{80}.

We regard the percentile of such importance that we feel justified in showing how it is calculated. We take the frequency distribution A from Table 1.

We shall demonstrate each step by calculating P_{80}, the 80th percentile, from Table 2.

The first step is to find the cumulative frequencies corresponding to each class interval. These are shown in column F, Table 2. The cumulative frequency (F) column is prepared by adding the successive class frequencies from the bottom to the top. The cumulative frequency corresponding to the interval 36-40 is found by adding the frequencies $1 + 2 + 3 + 4 = 10$ and gives the number of scores in the

distribution lying below the score 40·5. (Note.—The class interval 36-40 is regarded as extending from the lower limit 35·5 to the upper limit 40·5).

Similarly for class interval 41-45 the cumulative frequency is $1+2+3+4+5=15$.

To Calculate a Given Percentile from Grouped Data

TABLE 2

Calculation of Percentiles in a Frequency Distribution

Class Interval	Frequency	Cumulative Frequency	Percentage Cum. Freq.
(c)	(f)	(F)	(% F)
71-75	1	36	100
66-70	2	35	97.2
61-65	3	33	91.7
56-60	4	30	83.3
51-55	5	26	72.2
46-50	6	21	58.3
41-45	5	15	41.7
36-40	4	10	27.8
31-35	3	6	16.7
26-30	2	3	8.3
21-25	1	1	2.8
	N=36		

Any percentile may be calculated from the formula:—

$$P_X = l + \frac{(xN - F)}{f} \times c$$

where xN = Percentage of N the total frequency

l = lower limit of the class interval in which P_X lies

F = cumulative frequency of class interval immediately below l

f = number of scores within the class interval in which P_X falls

c = the size of the class interval.

To calculate P_{80}.

$$xN = \frac{80}{100} \times 36 = 28{\cdot}8$$

$\therefore$ P_{80} falls within class interval 56-60

$l = 55{\cdot}5$ $\qquad$ $f = 4$

$F = 26$ $\qquad$ $c = 5$

$$\therefore P_{80} = 55{\cdot}5 + \frac{(28{\cdot}8 - 26)}{4} \times 5$$
$$= 55{\cdot}5 + 3{\cdot}5$$
$$= 59$$

This means that 80% of the pupils in this examination scored marks below 59.

In some scaling procedures it is an advantage to look at percentiles and their corresponding scores the other way round and to find the percentile corresponding to a selected score, instead of the score for a selected percentile; for example, to find from grouped data the percentiles corresponding to the upper limits of class intervals 25·5, 30·5, 35·5 and so on. This can be very quickly done for the whole series of upper limits if a column of percentage cumulative frequencies is added to Table 2. From the Table we see that there are 21 scores out of a total of 36 below the value 50·5; stated differently 58·3 per cent. of scores lie below 50·5 or $50{\cdot}5 = P_{58{\cdot}3}$

(d) Standard Deviation

The most useful measure of the scatter of marks is the standard deviation. It is often denoted by the Greek letter σ. An example is perhaps the best way by which to illustrate what the standard deviation really is. Take the following marks scored by ten pupils, 84, 78, 72, 60, 56, 50, 49, 47, 43 and 41. The mean score is 58. The deviations of each score from the mean (the difference between each score and the mean) are 26, 20, 14, 2, -2, -8, -9, -11, -15 and -17. These deviations are squared to give 676, 400, 196, 4, 4, 64, 81, 121, 225 and 289. The sum of the squared deviations is 2,060 and therefore the mean squared deviation is 206. The square root of this is 14·3 and is the standard deviation. From the above example it can be seen that the standard deviation is the square root of the mean of the squares of the deviations from the mean.

The calculation of the standard deviation for a frequency distribution with any class interval is given in Appendix III.

The standard deviations for marks A, B and C are 12·1, 12·1 and 8·6 respectively showing that the scatter of marks given by

C

teachers A and B is the same and each is greater than that given by teacher C.

The Normal Frequency Distribution

We have shown how distributions of different kinds are most suitable for examinations of different types. There is one frequency distribution which is of great importance; it is called the normal frequency distribution. As an American has said, it is the lathe on which the statistician makes most of his tools. Take the following distribution of IQs:—

Table 3

Normal Distribution of IQs

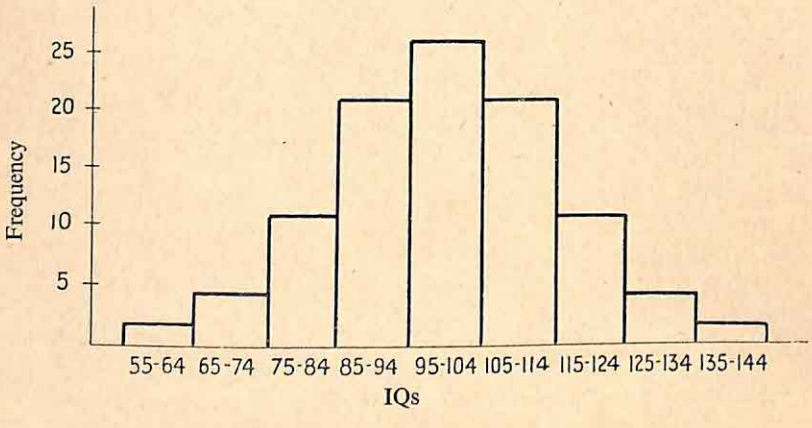
Class Interval	55-64	65-74	75-84	85-94	95-104	105-114	115-124	125-134	135-144
Frequency	1	4	11	21	26	21	11	4	1

The histogram of this distribution would be as shown in Fig. VI.

If a curve is drawn through the mid points of the top of each box, it will give a well-shaped symmetrical curve known as the normal curve. The curve is symmetrical about the mean of

Fig. VI

Histogram of Normal Distribution of IQs



the distribution. The second feature of the normal distribution is that the scores concentrate closely around the mean and taper off equally on either side; there is a slow decrease in the frequencies for a certain distance above and below the mean, a quicker decrease for a bit and then a slower decrease to the limit on both sides. (See Fig. VII).

Measuring from the Mean

Perhaps we should indicate at this point that the statistician finds it convenient to measure from the mean and not from zero. The reason for this is evident if we consider finding the mean of the following numbers—86, 89, 90, 92, 94, 95. By adding them and dividing the total, 546, by 6, we find that the average is 91. It is much simpler, however, to subtract 90 from each, transforming the numbers into -4, -1, 0, 2, 4, 5, giving a total of 6 and a mean of 1. When 90 is added again the mean would be $90+1=91$. This device is often used unconsciously. Golfers, for example, often count their scores by measuring from their probable average number of strokes for each hole; two above 4s for nine holes would mean a score of 38.

Another device adopted by the statistician is measuring from the mean in units of the standard deviation or sigma units, as they are called. For example, if the mean IQ of the group of children is 100 and the standard deviation is 15, an IQ of 130 would be $+2\sigma$ in sigma units because 130 is 30 above the mean, which is $+\frac{30}{15}\sigma$, that is $+2\sigma$ in terms of the standard deviation of 15. Similarly, an IQ of 80 is $-1{\cdot}33\sigma$ in sigma units.

In the diagram below, the distribution of IQs is shown in steps of 15; below the line are shown the corresponding sigma measurements from the mean 0.

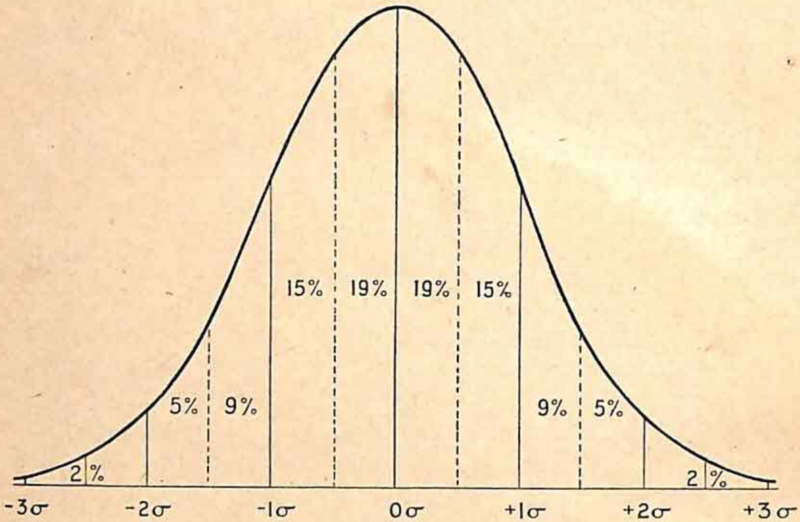
55	70	85	100	115	130	145
-3σ	-2σ	$-\sigma$	0	$+\sigma$	$+2\sigma$	$+3\sigma$

Figure VII shows how the area is divided up at $\frac{1}{2}\sigma$ intervals, correct to the nearest percentage.

Relationship between the Standard Deviation and the Area under the Normal Curve.

FIG. VII

Area under Normal Curve



For the purpose of this book it is convenient to think of the percentage of the total area as being cut off by an ordinate that moves from left to right. For example, the ordinate at -2σ cuts off approximately 2% of the area, the ordinate at -1σ cuts off 16% (2% + 5% + 9%) of the area, and so on. If we think of Fig. VII as representing the distribution of 100 scores in perfectly normal fashion, then there are two scores (out of 100) below -2σ. Another way of stating this is that the second percentile (P_2) corresponds to -2σ, P_{16} to -1σ, P_{50} to 0, P_{84} to $+1\sigma$, and so on.

If a normally distributed set of marks had a mean of 50 and a standard deviation of 10, then P_2 would correspond to 30 marks, P_{16} to 40, P_{50} to 50 and P_{84} to 60.

We have shown the relation between sigma scores and percentiles; for example, P_{84} corresponds to $+1\sigma$. The statistician has constructed tables giving all the percentiles

in terms of sigma units. One such table is given in Appendix IV from which the following figures are extracted:—

Standard Deviation	Percentile
−2.0	2.3
−1.5	6.7
−1.0	15.9
−0.5	30.8
0	50
+0.5	69.2
+1	84.1
+1.5	93.3
+2.0	97.7

There are some figures in this table to which special attention must be drawn: -1σ cuts off approximately 16% of the scores from the bottom, and $+1\sigma$ cuts off 16% of the scores from the top. If we can find the two scores that are 16% from the top (P_{84}) and from the bottom (P_{16}) respectively and subtract them, we get the range of scores that is equal to twice sigma.

For our benefit the statisticians have prepared another table which shows the sigmas and percentiles the other way round (Appendix V). From this table the following figures have been extracted:—

Percentile	Standard Deviation
1	−2.326
5	−1.645
10	−1.282
50	0.000
70	+0.524

If we know the mean and standard deviation of a normal distribution, we can calculate the scores corresponding to any percentile from Appendix V. For example, if in a normal distribution the mean is 50 and the standard deviation is 10, then

$$\begin{aligned} P_5 &= 50 - 10 \times 1{\cdot}645 \\ &= 33{\cdot}55 \end{aligned}$$

The reader is invited to check the following figures:—

$P_1 = 26{\cdot}74$	$P_{99} = 73{\cdot}26$
$P_5 = 33{\cdot}55$	$P_{95} = 66{\cdot}45$
$P_{10} = 37{\cdot}18$	$P_{90} = 62{\cdot}82$
$P_{20} = 41{\cdot}58$	$P_{80} = 58{\cdot}42$
$P_{30} = 44{\cdot}76$	$P_{70} = 55{\cdot}24$
$P_{40} = 47{\cdot}47$	$P_{60} = 52{\cdot}53$

$$P_{50} = 50{\cdot}00$$

If ordinary graph paper is used with the horizontal axis denoting percentiles and the vertical axis scores, we get a curve known as an *ogive*. (See Fig. VIII).

If specially ruled paper* known as Permille or Arithmetical Probability Paper is used the percentiles do not form a curve but a straight line.

To draw this line, of course, only two points are required. We need hardly point out the great saving in time and labour achieved by using Permille paper. To draw the normal line we might calculate P_{84} and P_{16} and draw the line joining these two points. (See Fig. IX).

When the distribution is not normal, percentiles calculated in the usual way and plotted on this special paper will not be on one line: the points will form when joined what is known as a zig-zag (see Fig. X).

The table on page 28 gives in percentile form the distribution of IQs in a secondary school.

The zig-zag is shown in Fig. X.

If the best fitting line is drawn among these points and the IQ values of P_{84} and P_{16} are read off, we have a good

* Data Sheet No. 37. *Arithmetic Probability*. (Wightman Mountain, Ltd., Artillery House, Artillery Row, Westminster, S.W.1).

Data Sheet No. 37. *Permille Paper*. (Hunt & Broadhurst, Ltd., Ideal Works, Botley Road, Oxford).

This is a form of graph paper in which the vertical lines are spaced, from the central line (P_{50}), at distances proportional to the sigma values of the percentiles (Appendix V). For example, P_{16} is -1σ from P_{50}, P_{90} is $+1{\cdot}28\sigma$, from P_{50}. The horizontal lines are equally spaced as in ordinary graph paper.

FIG. VIII

Percentile Curve or Ogive

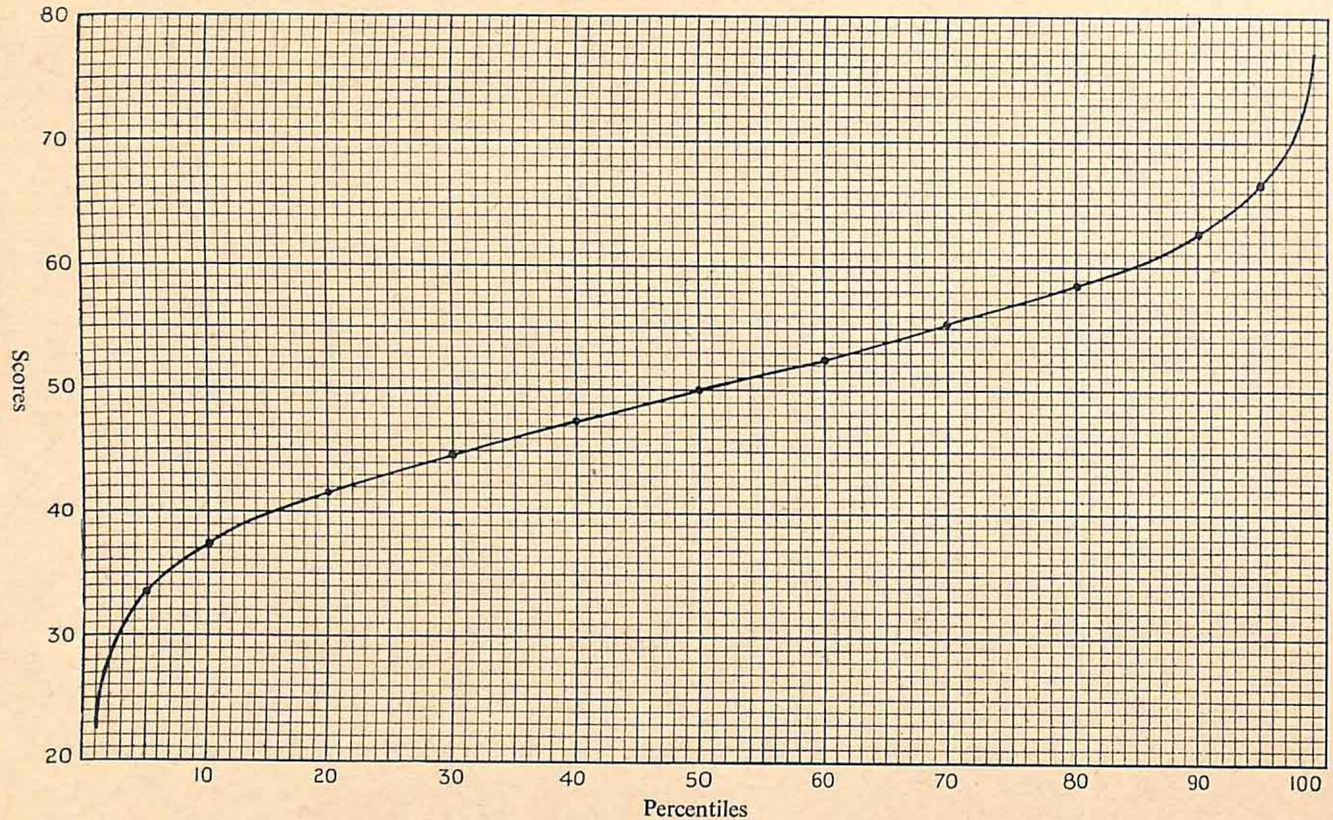


FIG. IX

Percentile Curve on Permille Paper

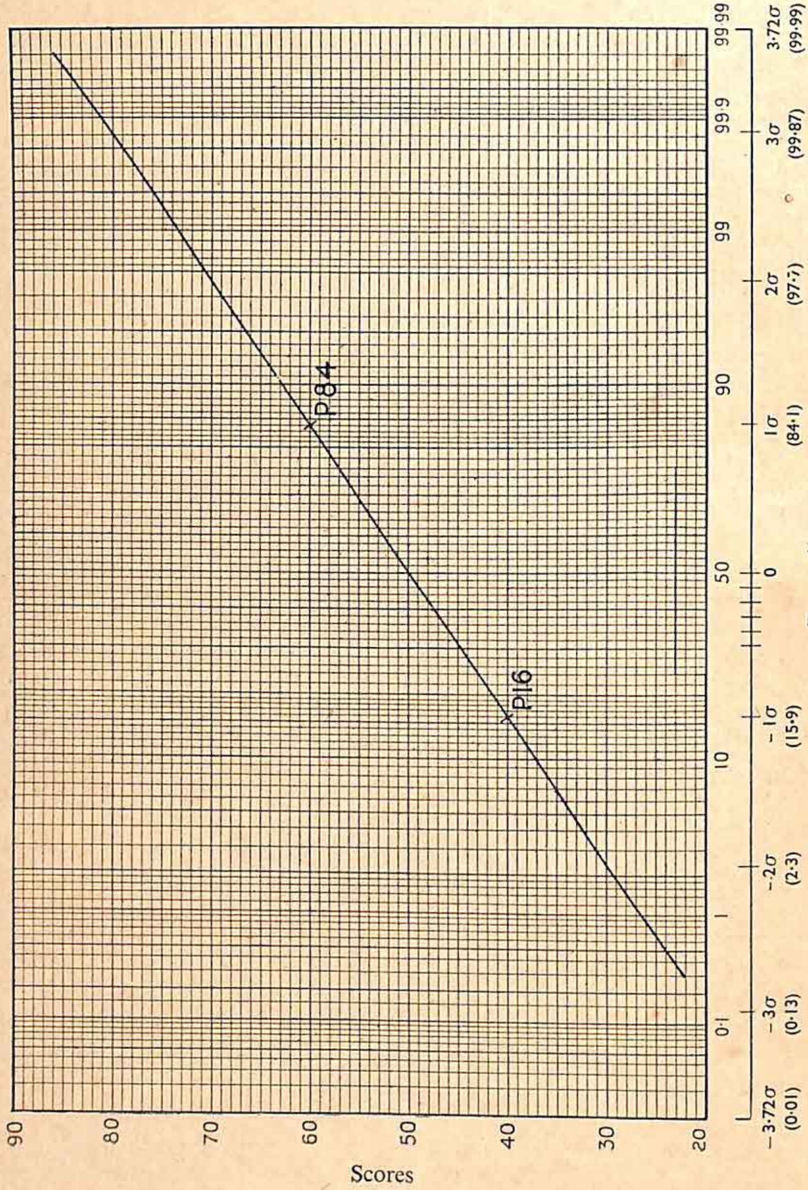
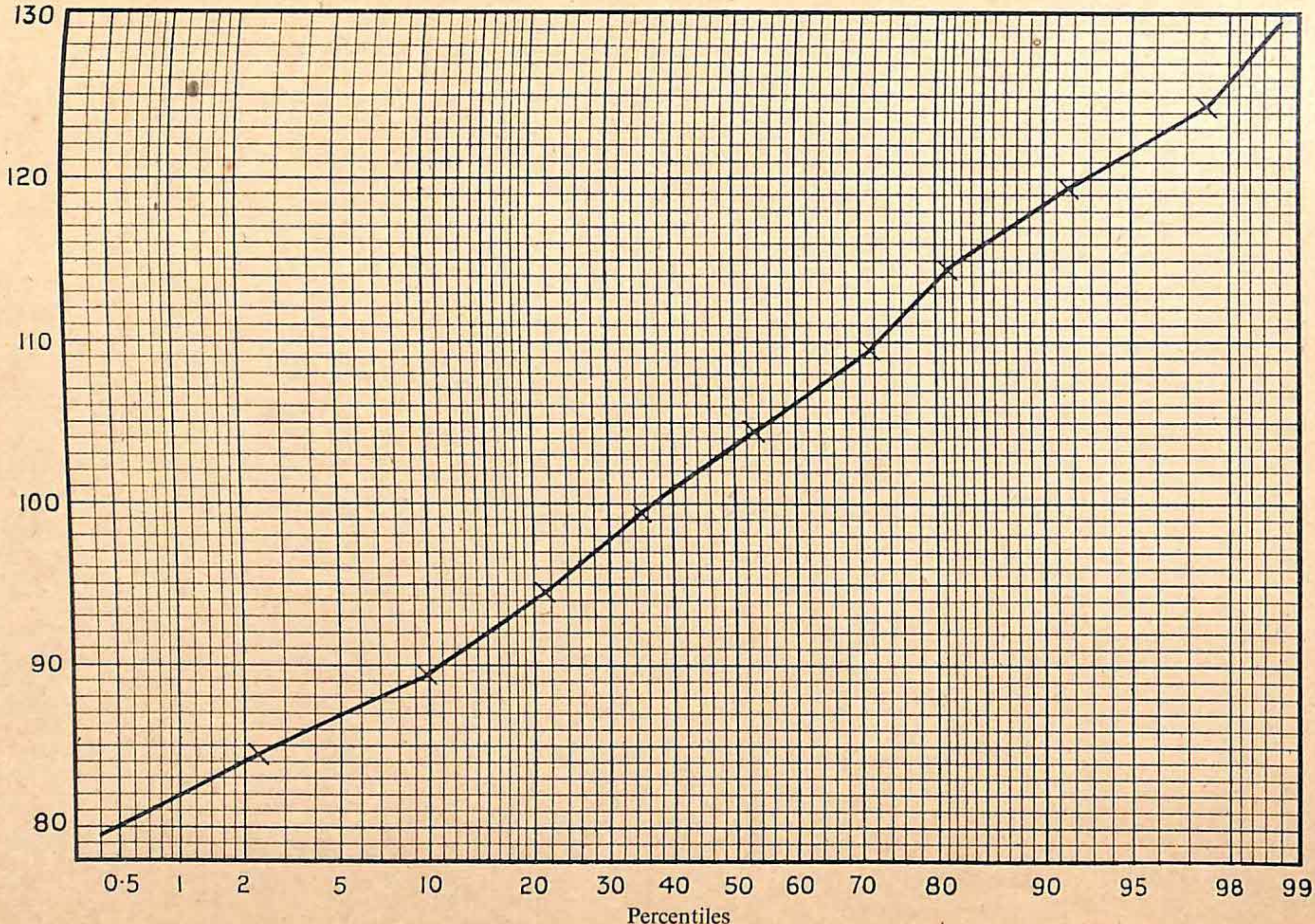


Fig. X

Graph of Percentile Distribution of IQs on Permille Paper



Percentile	IQ
0.4	79.5
2.3	84.5
9.8	89.5
21.4	94.5
35.3	99.5
53.1	104.5
71.8	109.5
81.3	114.5
91.9	119.5
97.5	124.5
98.8	129.5

approximation to the standard deviation of IQs in the school. In this case

$$P_{84} = 115$$
$$P_{16} = 95$$
$$P_{84} - P_{16} = 2\sigma = 20$$
$$\therefore \sigma = 10$$

The mean will be approximately $\dfrac{P_{84} + P_{16}}{2} = \tfrac{1}{2}(115 + 95)$

$$= \tfrac{1}{2} \times 210$$
$$= 105$$

This chapter will have attained its limited objective if it has put into the hands of the reader such of the elementary tools of the statistician as are needed for scaling and has shown him how to handle them.

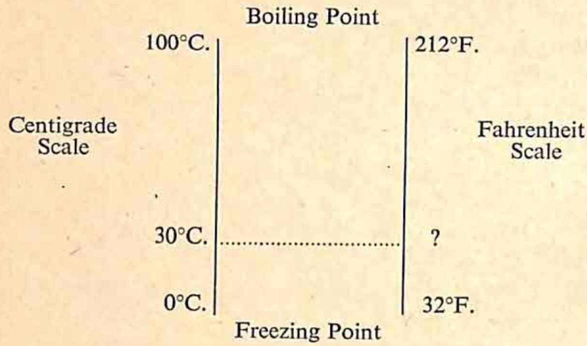
CHAPTER IV

FROM PHYSICAL TO MENTAL SCALING

SIR CYRIL BURT has stated the difficulties in interpreting examination marks as follows:—"When we turn from physical measurements to those of mental capacities or attainments a difficulty confronts us. In comparing physical magnitudes, we assume that all observers are using the same scale: if a French doctor takes a temperature with a Centigrade thermometer while his English colleague records the result in Fahrenheit, we make the necessary conversion before the two readings are compared. Similarly, in dealing with marks obtained in an examination we must also take into account possible differences of scale."* The two doctors know their respective zero-points and units. Two examiners, however, wishing to compare marks have no fixed points on their scales as the doctors have; they have to devise methods of inferring these from the whole run of the marks. Once they have found their zero-points and units they can convert in the same way as doctors do with temperatures. Let us digress for a little to see how the doctors might have done their conversions.

The quickest way would be to consult one of those thermometers that have the two scales lying on opposite sides of the mercury column. In the absence of such a thermometer or a wall chart or a ready made conversion table, they might fall back on the formula they had learnt at school. The formula forgotten, they would devise a solution in some such way as this: they would note (1) that the interval of temperature between the freezing point and boiling point of water is divided into 100 units on the Centigrade instrument and 180 on the Fahrenheit, and (2) that the effective zero-point of the Centigrade scale is 0 and the Fahrenheit 32.

* HARTOG, RHODES & BURT. *Marks of Examiners.* (Macmillan & Co., 1936).



If they wish to convert 30° Centigrade into the corresponding F degrees Fahrenheit they would write $\frac{F-32}{180}=\frac{30-0}{100}$ each expression being an equal fraction of the interval between the boiling point and the freezing point, the Big Unit of Temperature or B.U.T. as we may call it. By simple algebra:—

$$\therefore\ F-32=\frac{180}{100}\times 30$$

$$\therefore\ F=\frac{9}{5}\times 30+32$$

$$=86$$

In effect they would be using the well-known formula—

$$F=\frac{9}{5}C+32$$

where C is the temperature in degrees Centigrade. This formula is an instruction to the English doctor: to the zero of the Fahrenheit scale (32) add the Centigrade reading stretched out in the ratio 9:5.

The formula is all very well for the odd occasion; if several conversions have to be made at the same time a table or graph is much speedier.

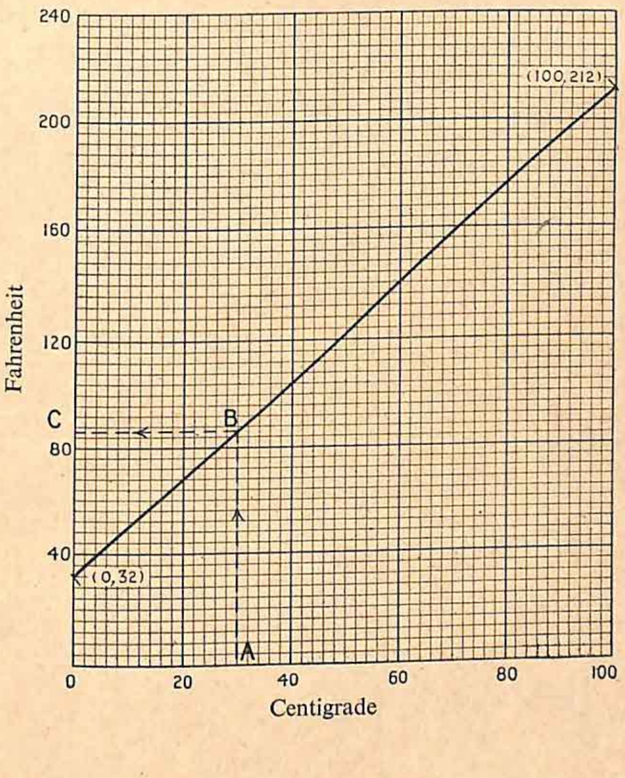
To construct a conversion table substitute in the formula, $F=\frac{9}{5}C+32$, values of Centigrade from 0 to 100.

Centigrade	Fahrenheit
0	32
1	33.8
2	35.6
⋮	⋮
100	212

The graph is constructed by drawing the straight line joining the points (0°C, 32°F) and (100°C, 212°F). To convert 30°C. to Fahrenheit, follow the vertical AB till it meets the graph at B and the horizontal line BC till it meets the vertical axis at C which gives the Fahrenheit reading.

FIG. XI

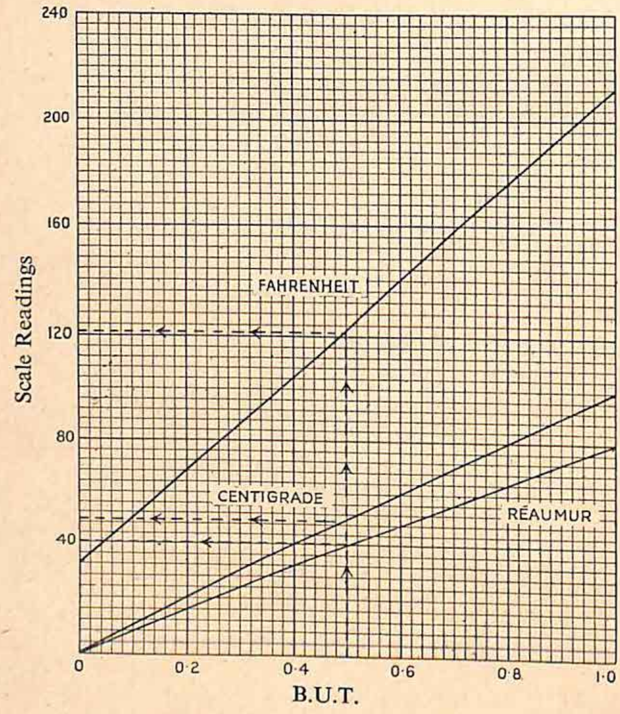
Converting Centigrade to Fahrenheit by Graphical Method



It would have complicated matters for the doctors if a Russian had been present with his Rēaumur scale (freezing point 0, boiling point 80). Figure XI enabled us to convert from one scale to another. The three doctors would have required three conversion graphs to compare their temperatures; obviously it would be much more convenient to compare all three at the same time. This can be done if there is some common scale against which each can be measured. There is a common scale at hand—that based on the interval between the boiling point and freezing point, the B.U.T. Figure XII below shows the conversion chart. Conversion can be made up any vertical line; for example,

FIG. XII

Conversion Chart for Fahrenheit, Centigrade and Reaumur Temperature



40°R. corresponds to 50°C. and to 122°F. Again 80°C. corresponds to 62°R. and 176°F.

Had a fourth doctor come along with yet another scale, freezing point 50 and boiling point 150, he could have been accommodated on our diagram and his readings scaled to that of any of his colleagues.

There is, however, a kind of thermometer that cannot be treated in this way. Let us imagine a "Hottentot" doctor who scales his own thermometer without taking into account the boiling point and freezing point of water. He simply divides his tube into 100 equal divisions, putting 0 just above the bulb and 100 near the top. His instrument will be quite useful up to a point; he can go round his hospital registering the temperature of each patient, and so put them in an "order of merit" of temperature. After finishing his round he can examine his layout or distribution of temperatures and come to the conclusion that certain patients have temperatures so much above the average, or so far below the average, that they require special attention. He would find possibly that, say, 56 was the average. To complete the illustration, let us suppose that the "Hottentot" doctor throws away his thermometer each night. With a new thermometer of a different bore and size of bulb, he will have to go through the same performance again, carefully examining his distribution of temperatures in order to single out each day the dangerous cases. His normal reading will be quite different except by chance.

The story of the "Hottentot" doctor is a near parable of the procedure which the teacher must adopt, willy nilly, in marking his papers. He may construct an examination paper in arithmetic with his "zero" at 0 and his "maximum" 100 and, just as the "Hottentot" doctor can put his patients in order of merit of temperature, the teacher can arrange his pupils in order of merit in arithmetic noting which have a high and which have a low achievement. He, too, might infer that 56 was a normal or average mark but a mark of 56 in itself will be meaningless to teachers of

any other subject. The next day he might set another examination in arithmetic. Again he will have to scrutinise the marks to find out the average mark. If the new average were found to be 56 it would be purely a chance occurence.

The English, French and Russian doctors could convert their readings because they had a common scale, the B.U.T., against which they could compare their temperatures. Teachers of arithmetic, English and mathematics in the same school can only compare their marks if there is some common scale or objective standard against which their marks can be measured. In the same way a Director of Education must have some objective standard with which he can compare, convert or scale the marks in arithmetic from different schools.

The educationist, like the scientist, must devise a formula, table or graph to convert or scale marks so that they have a meaning not only for himself but for others.

CHAPTER V

THE SCALING PROCEDURE

We have seen that an examination mark by itself has no meaning; the standard of marking and the scatter of the marks must be known, that is, the mean and the standard deviation of the distribution of the marks must be known.

The fundamental principle in scaling is to make the standard and scatter of marks in one set of marks the same as the standard and scatter in another set. For example, teachers' estimates in arithmetic from a number of schools may be scaled on an external test in arithmetic with the results that the mean and standard deviation of the estimates in each school become the same as the mean and standard deviation of the external test scores for that school.

No attempt will be made to establish the formula for scaling a series of marks but a parallel may be drawn with the conversion of Centigrade into Fahrenheit temperatures which will give the non-mathematical reader some idea of its derivation.

When a Fahrenheit thermometer and a Centigrade thermometer are placed in the same dish of water, they give different readings because they employ different scales. It was shown in Chapter IV that these two readings are related to each other by the equation.

$$\frac{F-32}{180}=\frac{C-0}{100}$$

$$\text{or } F=\frac{9}{5}C+32$$

This equation was derived from a knowledge of the fixed points and units of each scale.

In examination marks our fixed point is the class average and our unit the standard deviation of the class marks from the average. When marks are converted to this scale, we

can compare marks in different examinations, or we can compare examination marks with teachers' estimates.

Suppose, for example, that a pupil has scored 58 in an examination where the mean was 51 and the standard deviation 14. His standard score is $\frac{58-51}{14}=0{\cdot}5$. Suppose that in another examination applied to the class he scores 78, the average being 70 and the standard deviation 16. His standard score is $\frac{78-70}{16}=0{\cdot}5$. Our conclusion is that his performance is equally good in each examination.

To generalise the numerical argument, let us denote the mark obtained by our candidate in the first examination by X, the average mark by $\overline{X}$, and the standard deviation by σ_X; similarly Y, $\overline{Y}$ and σ_Y represent the mark, average mark and standard deviation in the second examination. Then the numerical relation

$$\frac{58-51}{14}=\frac{78-70}{16}$$

becomes $$\frac{X-\overline{X}}{\sigma_X}=\frac{Y-\overline{Y}}{\sigma_Y}$$

This equation tells us what mark X in the first examination corresponds to the mark Y in the second examination and *vice versa*. To find the mark X corresponding to a given Y we should write the equation in the form

$$X=\overline{X}+\frac{\sigma_X}{\sigma_Y}(Y-\overline{Y})$$

We can then calculate for each value of Y the corresponding value of X. This is called scaling Y on X, and the equation is called *the scaling equation.* $\frac{\sigma_X}{\sigma_Y}$ may be termed *the scaling ratio* or the *adjusting ratio.*

If this operation were carried out on the marks previously mentioned, where the average was 70 and the standard deviation 16, the new or "scaled" marks would have an

average of 51 and a standard deviation of 14. This is the essential point about scaling; it is an operation for changing the average and standard deviation of a set of marks to a new average and standard deviation.

One important use of this operation is in scaling teachers' estimates on the marks obtained by the class in an external examination. The need for scaling these estimates was pointed out in Chapter II. One way of scaling them is to give them the same average and standard deviation as the marks obtained by the class in an external examination in the subject for which the teachers' estimates were prepared. The Scaling Equation will be:—

$$T_S = M_E + \frac{\sigma_E}{\sigma_T}(T - M_T)$$

where T_S = Scaled teachers' estimate.
M_E = Mean test mark.
σ_E = Standard deviation of marks.
σ_T = Standard deviation of teachers' estimates.
T = Teachers' estimate.
M_T = Mean of teachers' estimates.

It will be obvious that the estimates for a class which has performed well in that examination will be adjusted to a corresponding level of excellence and that any tendency of the teacher to bunch estimates will be counteracted. A point that may not be clear at first sight, but which will

TABLE 4

Estimates and Marks in English for 20 Pupils

Pupil	Estimate	Examination Mark	Pupil	Estimate	Examination Mark
A	40	47	K	48	56
B	45	54	L	70	72
C	65	66	M	45	49
D	50	41	N	45	56
E	60	59	O	60	78
F	65	74	P	40	43
G	80	84	Q	60	66
H	35	40	R	35	32
I	42	49	S	40	41
J	45	54	T	60	65

become clearer when the following examples are studied, is that scaling does not alter the order of merit of the pupils as given by the teacher: the estimates are adjusted in size but the final order in the scaled estimates is exactly that of the original estimates.

There are given in Table 4 the teachers' estimates of attainment in English for each of the 20 pupils in a class, and the score made by that pupil in the external examination in English. These estimates will be scaled on the marks by two methods.

I. Arithmetical Method (Sigma Scaling)

From Table 4 above, the following are calculated.

(Note:—the full calculation is given in Appendix VI).

Average of teachers' estimates $= 51{\cdot}5$

Average of examination marks $= 56{\cdot}3$

$$\text{Scaling ratio} = \frac{\text{Standard deviation of marks}}{\text{Standard deviation of estimates}} = 1{\cdot}12$$

If the unscaled estimate is T, and the scaled estimate T_S, the scaling equation is

$$T_S = 56{\cdot}3 + 1{\cdot}12\,(T - 51{\cdot}5)$$

The application of this equation to the unscaled estimates will produce scaled estimates with an average of 56·3 and the same standard deviation as that of the marks.

The equation, when simplified, becomes:—

$$T_S = 1{\cdot}12T - 1{\cdot}4$$

On applying this equation to the estimates given above we obtain the scaled estimates shown in the fourth column of Table 5 below. For example, the first estimate 40 becomes $1{\cdot}12 \times 40 - 1{\cdot}4 = 43$ to the nearest whole number.

Possibility of negative scaled estimates

Suppose in another example the equation was:—

$$T_S = 1{\cdot}5T - 21$$

A teacher's estimate of 14 would give a scaled mark of 0 and all estimates below this would result in negative scaled estimates. For example, a teacher's estimate of 10 would give $T_S = -6$.

While a negative mark would not disturb the mathematician, the normal class teacher would hesitate to award a pupil a mark of −6 for English.

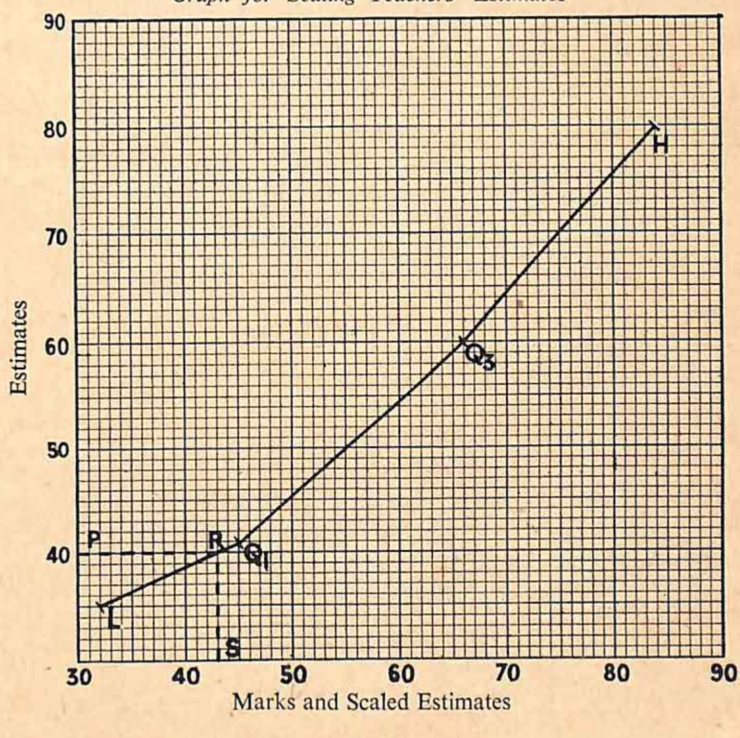
The occurrence of negative marks is one of the objections to the method of scaling by means of the scaling equation.

II. First Graphical Method.

McClelland in his "Selection for Secondary Education" gives a method which avoids the creation of negative marks. This system may best be illustrated by the example which we have used above with the scaling equation. The vertical axis denotes teachers' estimates and the horizontal axis the external examination marks. Four points are plotted:—

Fig. XIII

Graph for Scaling Teachers' Estimates



L—the lowest estimate (35) plotted against the lowest mark (32).

Q_1—the lower quartile of the estimates (41) plotted against the lower quartile of marks (45).

Q_3—the upper quartile of the estimates (60) plotted against the upper quartile of marks (66).

H—the highest estimate (80) plotted against the highest mark (84).

The line LQ_1Q_3H shows the approximate relation between estimates and marks for the class. The process of scaling consists in forcing the estimates into general agreement with the marks. The horizontal axis refers both to marks and to scaled estimates. An estimate P is scaled by following the line PR, RS to obtain the scaled estimate S.

The scaled estimates shown in the fifth column in Table 5 are obtained by this method.

TABLE 5

Estimates for English Scaled by Calculation and by Graphical Method

Pupil	Mark	Estimate	Scaled estimate	
			Calculation	Graph
A	47	40	43	43
B	54	45	49	49
C	66	65	71	70
D	41	50	55	55
E	59	60	66	66
F	74	65	71	70
G	84	80	88	84
H	40	35	38	32
I	49	42	46	46
J	54	45	49	49
K	56	48	52	53
L	72	70	77	75
M	49	45	49	49
N	56	45	49	49
O	78	60	66	66
P	43	40	43	43
Q	66	60	66	66
R	32	35	38	32
S	41	40	43	43
T	65	60	66	66

One disadvantage of this method is that an isolated very low—or very high—score or estimate exerts an unduly great influence on the scaled estimates in that quarter of the graph. The method is a rough form of a more precise method which will now be described.

III. Second Graphical Method (Percentile Scaling)

We use the following example to illustrate this more precise method of scaling.

An external test in English was set to several schools presenting candidates for the Senior Leaving Certificate. The results of this test were used to scale the teachers' estimates from the various schools. Table 6 gives the frequency distribution of the scores and estimates from school E.

Table 6

Frequency Distribution of Scores and Estimates in English in School E

Class Interval	Test Scores	Estimates
85-89	—	—
80-84	—	—
75-79	6	—
70-74	7	5
65-69	8	7
60-64	10	10
55-59	9	15
50-54	8	15
45-49	6	5
40-44	3	3
35-39	2	—
30-34	1	—
Total	60	60

As the test was set to all schools it gives a common standard against which to compare the estimates. The principle involved in this scaling procedure is that two scores of the same percentile standing are equivalent; for example, a pupil who has a Percentile Rank of 70 in the estimates is given, as his scaled estimate, the score of the pupil who has a Percentile Rank of 70 in the Test. In this

way the estimates are forced to conform to the same standard and scatter of marks as those of the scores made by pupils in that school. Corresponding scores are read off from a graphical picture whose construction we detail below.

TABLE 7

Calculation of Percentiles for Upper Limit of Class Intervals of Scores and Estimates in English in School E

	Test Scores			Estimates		
	1	2	3	4	5	6
Class Interval	f.	c.f.	% c.f.	f.	c.f.	% c.f.
75-79	6	60	100			
70-74	7	54	90.0	5	60	100
65-69	8	47	78.3	7	55	91.7
60-64	10	39	65.0	10	48	80.0
55-59	9	29	48.3	15	38	63.3
50-54	8	20	33.3	15	23	38.3
45-49	6	12	20.0	5	8	13.3
40-44	3	6	10.0	3	3	5.0
35-39	2	3	5.0			
30-34	1	1	1.7			

Step 1—Calculate the cumulative frequencies (c.f.) for test scores and teacher's estimates—columns 2 and 5, Table 7.

Step 2—Change these into percentage cumulative frequencies (% c.f.)—columns 3 and 6. (Columns 3 and 6 give the Percentile Rankings of the upper limit of scores in the interval).

Step 3—Draw a graph formed by plotting the points, percentile against test score (1·7, 34·5; 5·0, 39·5, etc.).

See Graph A. Fig. XIV.

Step 4—On the same sheet draw the graph formed by plotting percentiles against estimates (5, 44·5; 13·3, 49·5, etc.).

See Graph B. Fig. XIV.

Step 5—Convert teacher's estimates into scaled estimates by finding on Graph A the mark which has the same percentile standing as the estimate on Graph B. For example,

FIG. XIV

Graph for Scaling Teachers' Estimates in English in School E

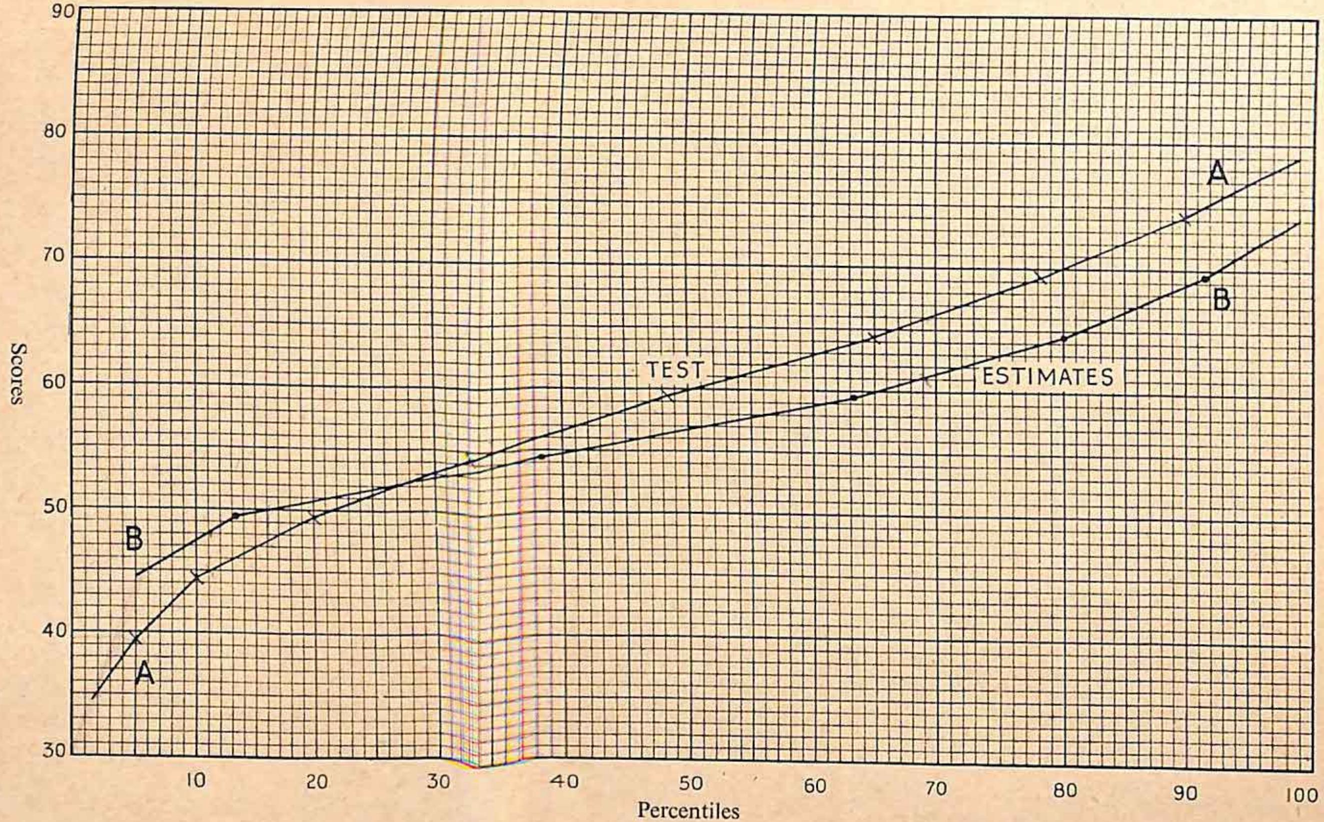
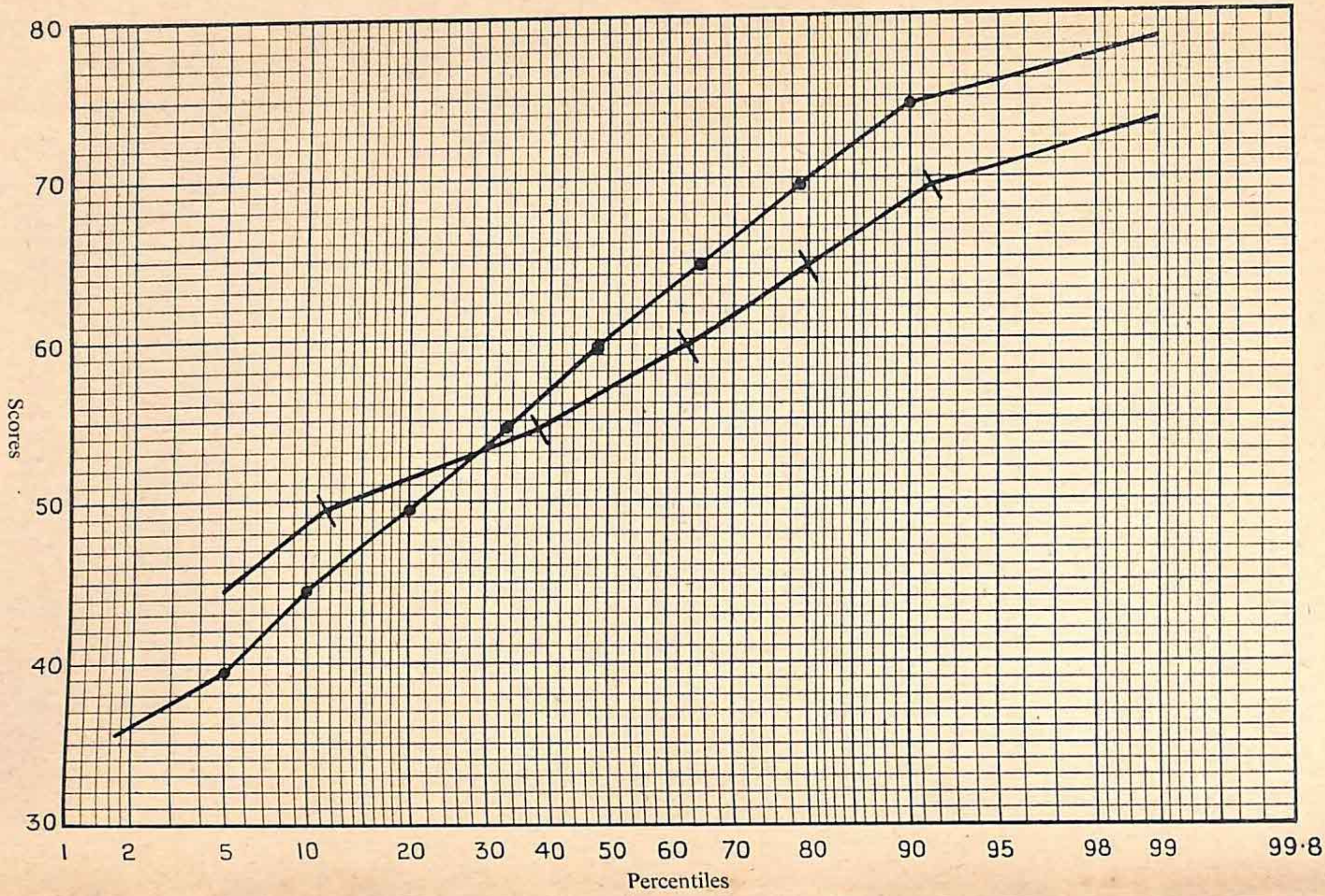


Fig. XV

Permille Graphs for Scaling Teachers' Estimates in English in School E



an estimate of 61 on Graph B and the corresponding point on Graph A each corresponds to a score of 66. Each of these marks has a percentile standing of 69.

Instead of using ordinary graph paper Permille paper might have been used (Fig. XV). It will be noted that almost identical results will be obtained.

In practice to save constant reference to the graphs and possible error, a conversion table would be constructed, showing the scaled estimate for each raw estimate. The scaled estimates are read off to the nearest whole number.

TABLE 8

Extract from Conversion Table for School E

Raw	Scaled
45	40
46	42
47	43
48	45
49	46
50	47
51	49
52	51
53	53
54	55
55	57
60	65
65	71
70	75

CHAPTER VI

SCALING TEACHERS' ESTIMATES—SOME FURTHER POINTS

In the previous chapter methods of scaling teachers' estimates were described. In this chapter it is proposed to discuss the problem in some of its wider aspects.

Perhaps the best known use of the scaling procedure is in schemes of transfer of pupils from primary to secondary education where the teachers are asked to submit estimates in English and arithmetic which are scaled on the results of an external test or examination in these same subjects. The Scottish Advisory Council on Education have outlined a scheme whereby the award of the Leaving Certificate may be based on teachers' scaled estimates.

Care must be taken that marks from the same school have the same significance. For example, a large school with three classes in English being presented for the Leaving Certificate will probably have three different teachers engaged in teaching these pupils. If all three teachers submit their estimates without any attempt at co-ordination the marks may as well be considered as coming from three different schools and must be scaled separately. If, as is desirable, a single scaling is to be made for the school one teacher or a team of teachers must be made responsible for co-ordinating the marks from the school.

Secondly, in selection for secondary education it has been found better to scale subject on subject; that is, estimates in English should be scaled on marks in English, and arithmetic estimates on arithmetic marks. It is economical of time and labour to scale the total estimate in English plus arithmetic on the total mark in these subjects, but the findings of the McClelland investigation were that the predictive power of the estimate was weakened by this merging of data.

The suggestion has been made that teachers' estimates in English and arithmetic might be scaled on intelligence

quotients. The attraction of this suggestion lies in the fact that it renders the external examination unnecessary for scaling purposes. The results of the investigation just mentioned show that a great deal of the predictive power of the scaled estimate is lost if this course is adopted. As will readily be seen, the method would not allow for differences in the attainment of a class in English and in arithmetic.

Thirdly, there is a limit placed on the feasibility of the scaling process by small numbers in a school. The scaling calculation previously described can be made when there are only a few candidates, but the assumptions underlying the procedure would no longer hold. The scaling methods described yield trustworthy results in general only when the number of pupils in the group is greater than sixteen or, better still, twenty. If the number is smaller, it may still be possible to scale successfully when the estimates and marks fulfil certain conditions, but this is a matter still open to experiment. In our Chapter XI on "Scaling in the Classroom," we show a method which gives at least a tentative solution of this problem.

It might be profitable to draw attention to the help that scaled estimates can give to a teacher. If estimates are scaled on standardised marks of the type to be described in Chapter VIII, and the scaled estimates are made available, as they ought to be, to the teacher concerned, he can at once see in what respect he has been successful or otherwise. For a teacher fresh to the work of the particular stage of primary or secondary education, this information might prove very valuable. There is the hope, too, that further practice, with knowledge of the results at each attempt, will enable teachers to prepare more accurate estimates each year.

The Difference Between Sigma and Percentile Scaling

It is with some reluctance that we introduce this topic as a full consideration of it would demand from the reader statistical knowledge beyond that given in the elementary outline in Chapter III; notwithstanding, we feel that, for completeness, it is necessary to refer to the matter briefly.

Readers who have closely followed the examples and who have possibly worked out examples of their own will have observed that the results obtained by different methods of scaling may not be exactly the same. In part, the differences are caused by the approximations made in reading graphs, and in rounding off figures in calculations, but there is a more fundamental difference between the sigma, or equation, method and the percentile method.

With the sigma method, the raw estimates or marks are made to conform with the marks in the external examination in two respects only, namely, in average and in scatter. Suppose, for example, that a set of estimates is distributed in either form shown on page 15 and that the external marks, on which they are to be scaled, are distributed normally. The use of sigma scaling will alter the average and scatter of the estimates but will preserve the lopsided character of the distribution. With the percentile method, the raw estimates or marks are made to agree with the external marks in *all* respects; the distribution of scaled estimates becomes identical with that of the external marks on which they are scaled. In the above example the use of percentile scaling will remove the lopsided nature of the estimates and produce scaled estimates which are normally distributed, as were the external marks, and have the same average and scatter as these marks.

The choice of scaling method is determined partly by the nature of the estimates or marks provided, and partly by the distribution of scaled estimates or marks desired. For the promotion stage, we favour the use of the sigma method on the grounds that teachers' estimates usually require adjustment only in standard and scatter, being distributed in passably normal fashion and that is the form in which we wish to have the scaled estimates. At the Leaving Certificate stage, on the other hand, the percentile method is recommended because the teachers' estimates are, for various reasons, far from normal and because there is a cluster about the 50 mark which must be removed to produce scaled estimates giving greater discrimination at that critical point.

CHAPTER VII

THE STANDARDISATION OF EXAMINATION MARKS IN PRACTICE

It has been shown that the distribution of examination marks should depend on the purpose of the examination. In examinations used for selection for secondary education it is preferable, though not essential, to have normally distributed marks. The word "normally" is here used in the sense of Chapter III. One reason is that these marks are often used along with intelligence quotients, and these, when the candidates are unselected and reasonably numerous, are usually distributed normally.

The marks obtained in examinations in English are frequently distributed almost in normal form, but in arithmetic it is sometimes found that the distributions are far from normal. It is possible to transform these "raw" marks into "standardised" marks which are normally distributed. What is more, the mean and standard deviations of the new marks can be adjusted to suit the examiner's requirements. This process of transforming raw marks, so that they (i) become normally distributed, (ii) have an agreed mean, and (iii) have an agreed standard deviation, is here called "standardisation."

In the example which follows, the agreed mean has been fixed at 65 and the standard deviation at 15. These figures have been chosen for the following reasons:—

(i) Intelligence quotients (as obtained from Moray House tests) are standardised to have a standard deviation of 15.

(ii) The number of scores above 100 in a normal distribution with this mean and standard deviation is very small; this fits in with the usual conception of 100 as the maximum of the scale. In practice there may be a few marks over 100.

(iii) The number of scores below 50 is about 16 per cent. of the total group, which again agrees roughly with the usual practice of teachers.

It must be made clear that standardisation is an operation to be applied to the scores of a whole county or city. The calculation which follows is one which would normally be made in an education office. The advantages of its application at a county level are several. It sets a standard which can be maintained year after year; a standardised mark of 60 in arithmetic in one year will mean exactly the same as 60 in the following year, provided that the number of candidates is reasonably large. A teacher who is asked to provide numerical estimates of attainment in English and in arithmetic has something on which to build; she knows or can be informed that an average pupil (considered on a county basis) should score 65, that only one pupil in 20 will score more than 95, and so on. In addition, since the IQ of the average child is 100, and the standard deviation of IQs is 15, while standardised marks have an average of 65 and a standard deviation of 15, there is a correspondence between standardised marks and the figures obtained by subtracting 35 (*i.e.*, 100 − 65) from IQs. It will be understood that this is only a rough guide, but it is at least of some assistance to a teacher to know that the arithmetic or English score appropriate to an IQ of 120 is 85 (*i.e.*, 120 − 35) while that for an IQ of 90 is 55.

Instead of attempting an explanation of the mathematical theory underlying the process, we shall illustrate by an example drawn from actual experience of standardisation. It is desired to standardise the scores of the 1,379 pupils whose marks are shown in Table 9 below. The desired mean is 65 and the standard deviation 15.

Example of Standardisation

Step 1—Prepare a table showing the frequencies of the raw marks grouped in class intervals of five. These class intervals[1] and frequencies are shown in columns 2 and 3 of Table 9.

[1] The more usual intervals are 0-4, 5-9, 10-14, etc.

Step 2—In column 1 enter the mark dividing each class interval from the one preceding it. This mark is denoted by the symbol x.

Step 3—Calculate the cumulative frequencies shown in column 4. Each gives the number of pupils scoring up to but not more than the corresponding value of x. The column is built up from the foot—70 = 48 + 22, 91 = 48 + 22 + 21, and so on.

TABLE 9

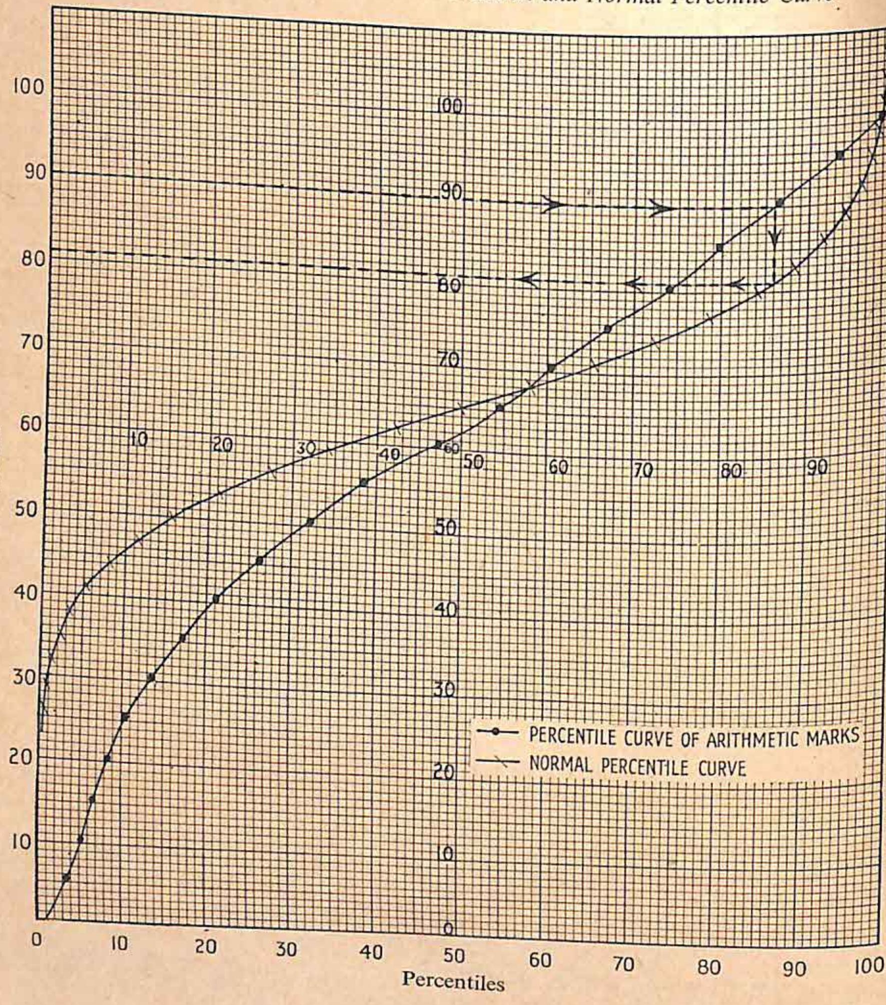
Distribution of Arithmetic Marks of 1,379 *Pupils*

(1) Boundary (x)	(2) Class Interval	(3) Frequency	(4) Cumulative Frequency	(5) Percentage
100.5	96-100	86	1379	100
95.5	91-95	99	1293	93.8
90.5	86-90	93	1194	86.6
85.5	81-85	79	1101	79.8
80.5	76-80	97	1022	74.1
75.5	71-75	90	925	67.1
70.5	66-70	86	835	60.6
65.5	61-65	98	749	54.3
60.5	56-60	122	651	47.2
55.5	51-55	83	529	38.4
50.5	46-50	86	446	32.3
45.5	41-45	70	360	26.1
40.5	36-40	53	290	21.0
35.5	31-35	45	237	17.2
30.5	26-30	46	192	13.9
25.5	21-25	32	146	10.6
20.5	16-20	23	114	8.3
15.5	11-15	21	91	6.6
10.5	6-10	22	70	5.1
5.5	0-5	48	48	3.5

Step 4—Convert the cumulative frequencies into percentages—column 5. The percentage of pupils who have scored less than 95·5 is now shown to be 93·8%, and so on. The last column therefore gives a percentile rating for each of the boundary marks listed in the first column.

Step 5—Draw a graph showing the relation between the marks and the percentile ratings. From this graph (shown in Figure XVI) it is possible to read off the percentile rating of any mark not shown in the table, and *vice versa*.

FIG. XVI
Percentile Curve of 1,379 Arithmetic Marks and Normal Percentile Curve



We must now prepare a similar graph for a normal distribution with mean 65 and standard deviation 15. This is done as follows:—

Step 6—Prepare the first column of Table 10 by writing down marks at intervals of three above and below the chosen mean of 65. The figure three is selected as being one-fifth of the standard deviation.

Step 7—For each mark enter the corresponding standard score, *i.e.*, the difference from the mean measured in units of standard deviation. For example, 68 is 3 marks or $0{\cdot}2\sigma$ above the mean and the standard score is therefore $+0{\cdot}2\sigma$; 35 is 30 marks or 2σ below the mean, and the standard score is -2σ.

TABLE 10

Percentiles for Normal Distribution with Mean 65 *and Standard Deviation* 15

Mark	Standard Score	Percentile	Mark	Standard Score	Percentile
104	2.6	99.5	62	−0.2	42.1
101	2.4	99.2	59	−0.4	34.5
98	2.2	98.6	56	−0.6	27.4
95	2.0	97.7	53	−0.8	21.2
92	1.8	96.4	50	−1.0	15.9
89	1.6	94.5	47	−1.2	11.5
86	1.4	91.9	44	−1.4	8.1
83	1.2	88.5	41	−1.6	5.5
80	1.0	84.1	38	−1.8	3.6
77	0.8	78.8	35	−2.0	2.3
74	0.6	72.6	32	−2.2	1.4
71	0.4	65.5	29	−2.4	0.8
68	0.2	57.9	26	−2.6	0.5
65	0.0	50.0	23	−2.8	0.3

Step 8—From Appendix IV read off the percentile rating corresponding to each standard score. This is entered in the third column.

Step 9—Graph these percentile ratings opposite the corresponding marks on the graph previously drawn (Fig. XVI).

For each percentile rating there are now two marks, the first obtained from the original set of marks, and the second from the normal distribution. The second mark is the standardised mark corresponding to the first or raw mark. For example, a raw mark of 90 becomes a standardised mark of 81, both having the percentile rating 86. The process of

reading off the standardised marks in this case is shown diagrammatically by the arrows.

Step 10—To save time and possible inaccuracy in repeated reading of the graphs, prepare from them a standardisation table (Table 11) showing for each raw mark the standardised mark.

It should be noted that although this calculation was based on grouped scores, the final table operates for each mark

TABLE 11

For Converting Raw Marks in Arithmetic into Standardised Marks

Raw Mark	Std. Mark	Raw Mark	Std. Mark	Raw Mark	Std. Mark	Raw Mark	Std. Mark	Raw Mark	Std. Mark
100	100	80	74	60	63	40	53	20	44
99	95	79	74	59	63	39	52	19	43
98	92	78	73	58	62	38	52	18	43
97	90	77	72	57	62	37	52	17	43
96	88	76	72	56	61	36	51	16	42
95	86	75	71	55	61	35	51	15	42
94	85	74	71	54	60	34	50	14	42
93	84	73	70	53	60	33	49	13	42
92	83	72	70	52	59	32	49	12	41
91	82	71	69	51	58	31	48	11	41
90	81	70	69	50	58	30	48	10	41
89	80	69	68	49	58	29	48	9	41
88	80	68	68	48	57	28	47	8	40
87	79	67	67	47	57	27	47	7	39
86	78	66	67	46	56	26	46	6	38
85	77	65	66	45	56	25	46	5	38
84	77	64	66	44	55	24	45	4	37
83	65	63	65	43	55	23	45	3	35
82	75	62	65	42	54	22	45	2	33
81	75	61	64	41	53	21	44	1	31
								0	29

from 0 to 100. The raw score of each of the 1,379 candidates sitting the arithmetic examination can now be transformed to the standardised score.

STANDARDISATION USING PERMILLE PAPER

The work of standardisation can be greatly eased by the use of Permille paper. This paper, as already stated, is so constructed that the curve of percentile ratings for the normally distributed marks becomes a straight line. Table 10 is no longer required in its entirety; only two points are required

for the straight line. If the reader cares to plot the remaining points in Table 10 on this chart he will find that the points fall on the straight line. A further advantage is that the conversion at the upper and lower extremes of the scale becomes more accurate and much easier.

The first four steps of the standardisation process remain as before, but the work is again made easier at the ends of the scale. The subsequent steps are:—

(5) Graph the percentile ratings against their corresponding marks, using Permille paper. Join the points by a series of straight lines or by a smooth curve. Note that on this paper the 100th and 0th percentiles are off the scale. (Figure XVII).

(6) Calculate, using Appendix IV, the two percentile ratings corresponding to the marks one σ unit above and one σ unit below the mean. For example, a mark 80 corresponds to $+1\sigma$ with a percentile rating 84·1 and a mark 50 corresponds to -1σ and percentile rating 15·9.

(7) Plot these two ratings against their marks on the Permille paper. Draw the straight line passing through the two points, extending it at each end.

The standardised marks may now be read from the graph as before, and a standardisation table constructed. The reader will observe that Figures XVI and XVII yield practically the same standardisation table.

In practice the standardisation of examination marks is most readily done by using Permille paper and we strongly recommend its adoption.

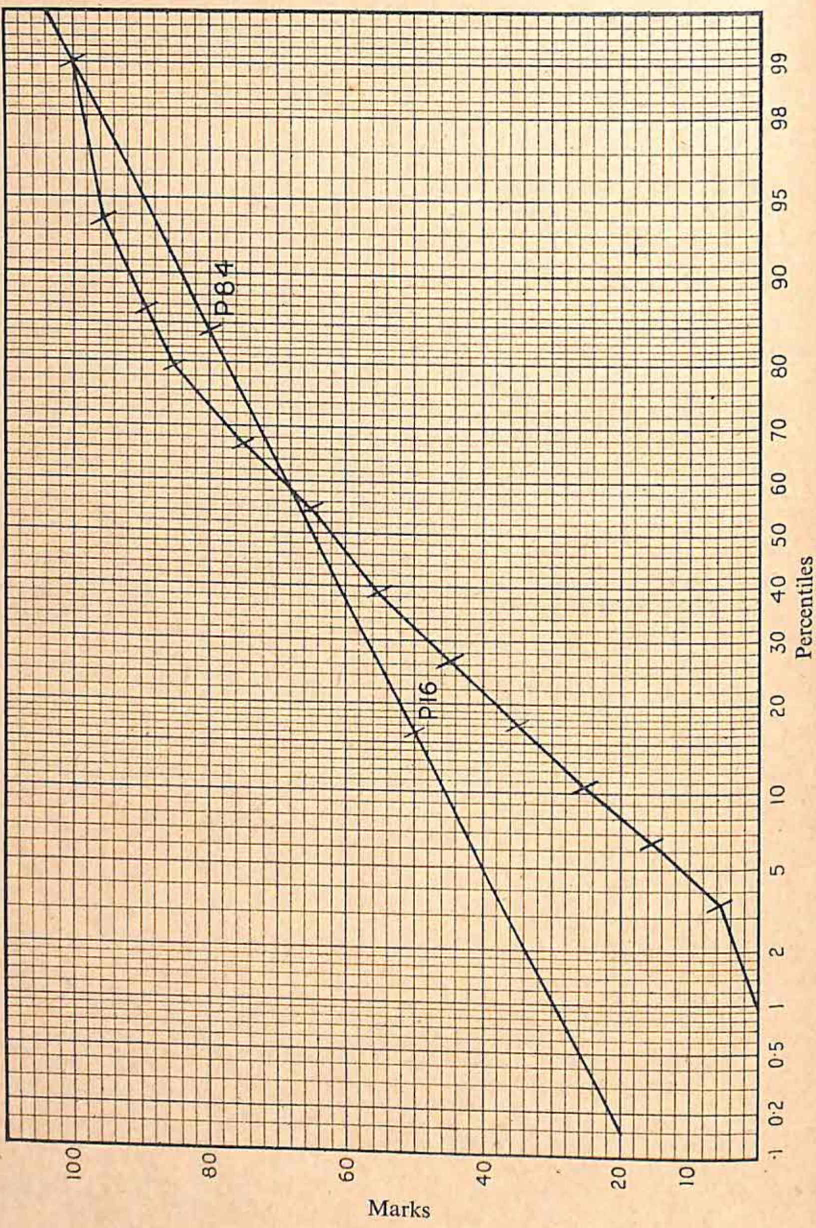


FIG. XVII

Percentile Line of 1379 Arithmetic Marks and Normal Percentile Line

($M=65$, $\sigma=15$)

CHAPTER VIII

PROMOTION FROM PRIMARY TO SECONDARY EDUCATION

THE Advisory Council in their Report on Primary Education outline a scheme for the transfer of pupils from primary to secondary education. It is recommended that the main basis for selection should be two intelligence tests, standardised attainment tests in English and arithmetic, and teachers' estimates of attainment in English and arithmetic. This set of six scores is sometimes referred to as the "battery" of tests.

The teachers' estimates must be "properly scaled so that the standard from school to school may be comparable."

The final mark of each pupil is found by taking the average of his scores in the intelligence tests, attainment tests and of the teacher's scaled estimates. This mark is known as the summary mark.

THE WEIGHTING OF TESTS IN THE PROMOTION BATTERY

If the scores in these tests are added in their "raw" state, each test has a weight proportional to its standard deviation. Thus, if the English and arithmetic marks have standard deviations of 10 and 15 respectively, the arithmetic mark has one and a half times the weight of the English mark when these are combined. McClelland found in his investigation that it was impossible to state exactly the weight that should be given to each of the tests in the battery. He recommended that equal weight should be given to each. In practice, not a great deal of difference is made if the raw marks of the tests are used, provided that the standard deviations of the tests are not too greatly different, but the following method gives the tests exactly equal weight.

Intelligence tests of the Moray House type are constructed to give a standard deviation of 15 for an age group. If the tests in English and in arithmetic are each standardised, by

the process described in Chapter VII, to have the same standard deviation of 15, then the teachers' estimates, which are scaled on the examination marks, will also have a standard deviation of 15 in each of these subjects. All six measures have now the same standard deviation, and carry equal weight when they are added.

Outline of a Promotion Scheme

A promotion scheme making use of the methods outlined above incorporates the following processes.

(a) The administration of two intelligence tests to each pupil.

The intelligence quotients can be obtained by testing each year a two year group, including pupils in Primary IV and Primary V. This has the additional advantage that any indisposition of the pupil is not likely to affect more than one test.

(b) The administration of tests in English and in arithmetic to each pupil.

The administration of the tests in English and arithmetic is a normal feature of Scottish educational procedure and does not need further explanation.

(c) The standardising of the examination marks in English and arithmetic.

The method of standardising these marks has been explained in Chapter VII.

(d) The scaling of teachers' estimates in English and arithmetic.

We favour the sigma method of scaling teachers' estimates which has been described in Chapter V.

(e) The calculation of the pass mark.

The average of the six measures is the summary mark for the given pupil. These summary marks form a distribution with a mean and standard deviation which can be calculated. The recommendation of the McClelland Report is that the pass mark for admission to a senior secondary course should be fixed at the summary mark which is 0·70 times the standard

deviation above the mean, and this mark can therefore be calculated.* A second and more tentative recommendation is that the borderline between the ordinary secondary course and a modified course for the weaker pupils should be about one standard deviation below the general mean. This mark can also be calculated from the summary marks.

THE OPERATION OF A PROMOTION SCHEME

While it is easy to indicate in a few sentences how each step is carried out, the operation of the scheme in a city or county is, of course, a much more laborious business. It might therefore be interesting to give a more detailed description of the process actually employed in a county operating the scheme.

For each pupil, there was prepared in Primary IV a card on which were entered the usual particulars, *e.g.*, name, date of birth, school. After the first intelligence test the intelligence quotient (IQ_1) was recorded on this card; similarly the second intelligence quotient (IQ_2) when obtained was also entered. For both of these entries the headteacher was responsible.

Prior to the examination, the headteacher entered on each card the teachers' estimates in English and in arithmetic. These were placed in the boxes marked (a) and (b) in Figure XVIII under the heading "raw." The cards were then sent to the Education Office where the estimates were recorded in summary form, and preliminary steps taken to scale the estimates.

When the cards were returned to the schools and the examinations had been held, the marks obtained were entered in boxes (c) and (d), under the heading "raw." The headteacher entered on the card his final recommendation regarding a suitable course, the parents' choice, and any notes, and the cards were returned to the office.

* This recommendation was based on the results obtained in a city, and on the courses being provided there. The Report goes on to suggest that further inquiry should be made into suitable passmarks for different types of area and different courses.

From the distribution of marks obtained from the cards, the office staff standardised the marks in English and arithmetic to a mean of 65 and a standard deviation of 15 in each subject and entered the standardised marks in English (Qe) and arithmetic (Qa) in boxes (e) and (f). The teachers' estimates were then scaled on the standardised marks, and the scaled estimates in English (Tes) and arithmetic (Tas)

FIG. XVIII
Record of Qualifying Data

		Raw	Battery	
Intelligence	Test I.			IQ_1
	Test II.			IQ_2
English	Test	c	e	Qe
	Estimate	a	g	Tes
Arithmetic	Test	d	f	Qa
	Estimate	b	h	Tas
Adjusted Average				

entered in boxes (g) and (h). These processes were also executed in the education office.

It will be observed that the six measures IQ_1, IQ_2, Qe, Tes, Qa, Tas, thus lay in a column on the edge of the card making summation easy, and facilitating also a comparison of the various measures so that discrepancies might be investigated.

The average of the IQs was 100 and that of the remaining marks was 65. If the six measures were added and the total

divided by six, the resulting mark would lose some of the significance that each of the components had. If a figure of 70, *i.e.*, $2(100-65)$ were subtracted from the total before division, then the average of the summary marks so adjusted was 65. This procedure of subtracting 70 before division was therefore adopted, and the resulting summary mark is the "adjusted average" shown in Fig. XVIII.

A similar argument and process could not be applied to standard deviations. Each of the six measures had a standard deviation of 15, but the standard deviation of the summary marks obtained from them need not have that value, in fact it is always less. In one year in which this scheme was applied, the standard deviation of the summary marks was approximately 13. The upper pass mark was therefore $65+0{\cdot}70\times 13=74$, and the lower pass mark $65-1\times 13=52$. These figures are only approximate, and are quoted to demonstrate that the pass marks obtained by this method have a general similarity to the percentage marks of the class teacher.

Borderline Cases

It might be well to emphasise here that this battery of tests, although the best found in the McClelland investigation, does not have perfect predictive power. Its findings must not be rigidly applied, and borderline cases must receive special consideration. The recommendation of the headteacher, the wish of the parent and other factors must be taken into account before a final decision is made.

The next step in the scheme described in this section was therefore a study of the cases where the recommendations of the battery, of the headteacher and of the parent were not in agreement. Although those considering these cases were able to formulate some general principles, these were of an experimental nature, and are therefore not stated here. This field is one which is ripe for experiment.

CHAPTER IX

THE LEAVING CERTIFICATE STAGE

THE awarding of a certificate to pupils leaving school has occupied the minds of educationists both in England and in Scotland in recent years. Attacks on the nature of the School Certificate in England and the Senior Leaving Certificate in Scotland and the method of awarding them have not been uncommon. We do not wish to enter the lists either for or against the method of awarding these certificates but it is clear that any method which relies solely upon the results of an external examination is bound to be unjust to many pupils. It is equally clear that teachers' estimates, unless they have been scaled, are just as unsatisfactory. We feel strongly, however, that the teacher's order of merit is the one which must be accepted and must therefore be taken into account in the award of any such certificate.

As our experience has been in Scottish schools we have drawn our illustrations from the Senior Leaving Certificate but the principles we lay down are applicable to any external examination.

The Advisory Council on Education in Scotland in their report on Secondary Education have suggested new principles governing the award of the Leaving Certificate and have outlined a method of examination which involves the use of the scaling procedure. A clear distinction must be made between the nature of the certificate and the nature of the examination. We here are concerned, not with the recommendations regarding the nature of the certificate, but only with the method of examination suggested.

Briefly, the proposal is to award the certificate on the teachers' estimates which would be scaled on the results of objective tests. The estimates in any subject would be scaled on the results of the objective test in the same subject and the scaling procedure would be carried out school by school.

The fundamental assumption underlying this proposal is that the teacher's order of merit is most likely to be the correct one. Educationists generally will agree that after four or five years' experience, a teacher should be able to assess a pupil's attainment in a subject much more accurately than does a single written external examination. The scaling procedure is introduced only to ensure that a uniform standard of marking is maintained between schools. Scaling does not alter the teacher's ranking of his pupils.

TABLE 12.

The Raw and Scaled Estimates of 13 Pupils Selected from Schools A and B

Scaled Estimate School A	Estimate	Scaled Estimate School B
53	47	37
55	48	38
56	49	40
60	50	44
63	51	48
64	52	52
65	53	55
66	54	57
68	56	61
73	58	65
77	60	69
81	65	76
83	70	82

In Table 12 we give the teachers' estimates in English of a number of pupils selected from two schools along with the corresponding adjusted estimates scaled on the results of an objective test in English. The scaled estimates did not alter the order of merit although there is an alteration in the size of the marks.

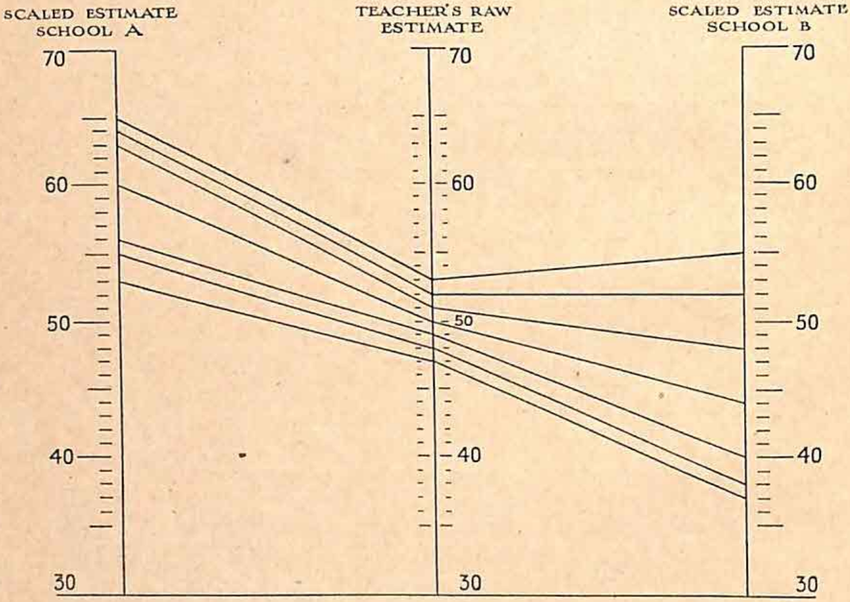
For the Senior Leaving Certificate teachers are asked to give a mark of 50 to a pupil who will just pass. This estimate of 50 is therefore a critical mark.

Let us look more closely at what happens in schools A and B to the estimates round about 50. The following diagram illustrates what happens to estimates from 47 to 53 in each school.

Scaling has produced some striking results; for example, a pupil estimated at 47 in school A has a scaled mark of 53 while a pupil with this same estimate in school B has a scaled mark of 37. The estimated pass mark of 50 becomes 60 in School A and 44 in school B. Teachers will find it difficult to believe that such different standards of marking occur.

FIG. XIX

Results of Scaling Leaving Certificate Estimates from Schools A and B



The diagram also shows that the order of pupils does not alter. For the group of pupils with marks between 47 and 53, the pupils with 47 in both schools still have the lowest marks while the pupils with 53 have the highest scaled marks.

It must also be clear that a pupil's mark in the external test does not determine whether that pupil "passes" or "fails." For example, in school A the pupil estimated at 58 scored a low mark, 42, in the test but this test score has no effect on the pupil's rank. It is only used in the calculation

of the mean and standard deviation of the test scores for school A. The teacher's order of merit or ranking remains unchanged and when the pass mark is fixed, all pupils above that mark pass and those below fail. Again in school A, the pupil estimated at 48 scored a relatively high mark, 50, in the test but this does not "boost" his scaled estimate.

This method of taking the teacher's order of ranking as the correct one and adjusting his standard of marking has much to recommend it. There seems little justification, for example, for the cases which occurred in a school where a pupil estimated at 62 failed and another estimated at 38 passed. Such utter disregard for the teacher's judgment is enough to bring any system of examination into disrepute. Under no circumstances should a certificate be awarded solely on the result of a single written examination.

THE PASS MARK

The Advisory Council suggest two methods of fixing the pass mark :—

(a) by a group of inspectors or examiners making a careful study of the results of the external test and the teachers' estimates for a number of representative classes;

(b) by fixing the pass mark at a certain multiple (to be determined by experiment) of the standard deviation of the marks of all candidates in the external test from the mean of the marks of all candidates in the external test.

The first method is subjective, depending to a large extent on the examiners' attitude. Research has shown clearly that there can be no guarantee that the standard fixed will be uniform from year to year by such a method.

The experimental methods of fixing the pass mark, while possessing the attributes of objectivity, are by no means easy to work out in practice. In selecting pupils for secondary education it is possible to carry out a follow-up investigation and compare the examination marks at the primary stage with the actual achievement in the secondary school. By such a method it is possible to find the mark which must be scored

before a pupil has a fifty-fifty chance of success in the secondary school. This is not possible at the Leaving Certificate stage as the pupils leave school and there is no obvious criterion of success.

The pass mark may be fixed by determining in advance the percentage of pupils who are to pass. It is then an easy matter to find the mark which will cut off the fixed percentage of successful candidates. Objection to this procedure may be taken on the grounds that the calibre of the pupils may vary from year to year. This is not likely to occur when the sample of candidates is drawn from schools over the whole country.

Technical methods of fixing the pass mark can be justly criticised but it must be remembered that the non-technical methods can be subjected to the same criticisms and several others besides.

Objective Tests at the Leaving Certificate Stage

Many criticisms have been made of the suggestion that objective tests should be used at the Leaving Certificate stage. These criticisms have been answered in the Advisory Council's report and there is no need to repeat them here. In Appendix I we have given examples of experimental types of these tests. While it is agreed that objective tests do not measure the power to marshall ideas or to assess literature, we suggest that the written examination is at best a poor measuring rod of these abilities.

Another objection to the use of objective tests is that teachers would coach for them and the broader aspects of education would be lost. While it is admitted that the poor teacher would be misled into attempting to raise the school mean to obtain a large number of passes, we have confidence in the profession as a whole that they would still retain their ideals and vision. In the second place, with greater experience in the construction of these tests, it will probably become more and more difficult to coach successfully for them; the full possibilities of the objective tests have not yet been realised.

Scaling Leaving Certificate Teachers' Estimates

An experiment was conducted in a number of senior secondary schools in which all pupils sat an objective test in the subjects for which they were presented for the Leaving Certificate. We demonstrate the scaling technique from the results in English in one of the schools.

The technique adopted is essentially the percentile method shown in Chapter V. The steps in summarised form are as follows:—

Step 1—Find the frequency distribution of the teachers' estimates and objective test scores.

Step 2—Calculate the percentiles for the upper limit of the class intervals. If the scatter of the estimates is not sufficiently great, additional percentiles may have to be calculated to give the line of percentiles.

Step 3—On Permille paper draw the line of percentiles of the objective test scores.

(See Graph A, Figure XX).

Step 4—On the same sheet draw the line of percentiles of the teachers' estimates.

(See Graph B, Figure XX).

Step 5—Convert teachers' raw estimates into scaled estimates by finding on Graph A the marks which have the same percentile standing as the estimates on Graph B. For example, an estimate of 60 falls on Graph B at L and the corresponding point on Graph A is at M which corresponds to a score of 71. Each of these marks has a percentile standing of 80.

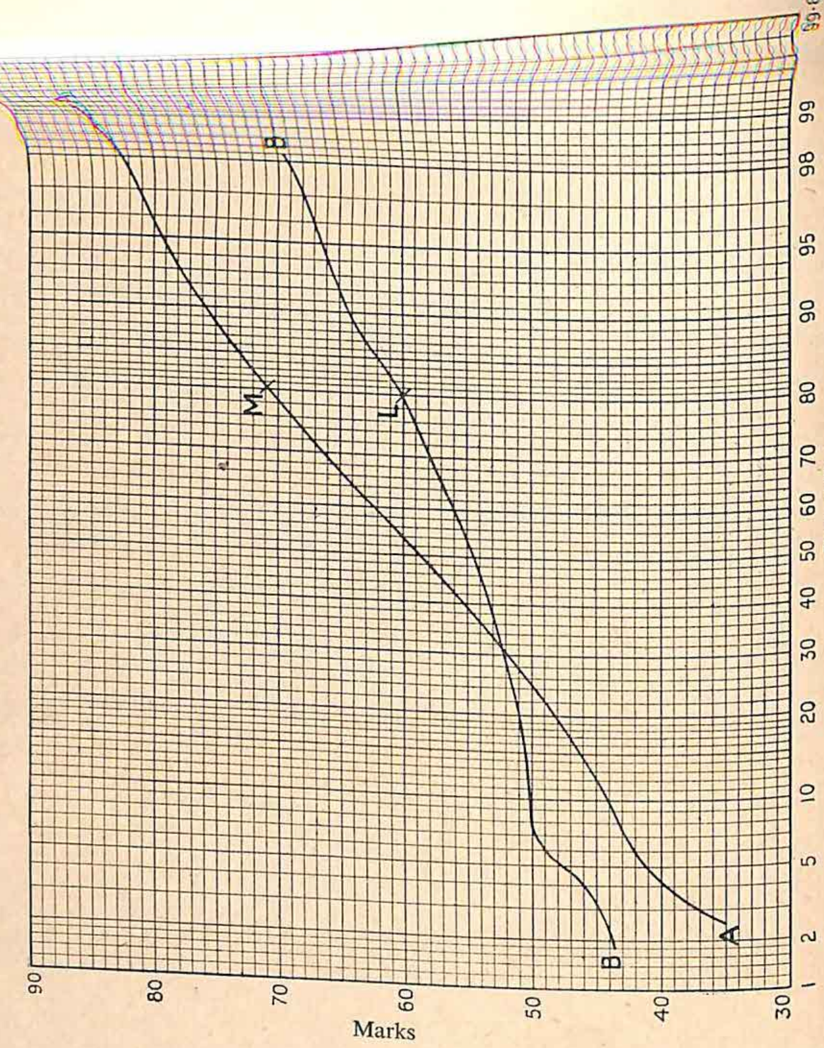
Step 6—Prepare a conversion table from which can be read off any scaled estimate corresponding to any raw estimate.

Small Groups

Perhaps the most serious criticism of the scaling procedure is that it cannot be used when the number in the group is small. It must be borne in mind that at the Leaving Certificate stage in small schools and in subjects such as Latin, the number of candidates presented is generally few.

FIG. XX

Graphs for Scaling Leaving Certificate Teachers' Estimates



In Chapter XI we show how the scaling technique can be used with groups down to ten pupils.

Where the number of cases is less than ten, the teacher's standard of marking can be assessed to some extent by a careful study of the estimates, the test scores and the data on which the teachers have based their estimates. Our experience has shown that some basis for adjusting marks is possible when only a few candidates are under consideration. The relation between estimates and marks in the external test is often apparent when the corresponding estimates and marks are plotted on graph paper. Even with four or five marks, the examiner who has acquired an intimate knowledge of how the scaling operation has worked over a large number of schools, can often give a good approximation to the alteration which must be made on the estimates. We do not propose to discuss this point in detail as our researches are still incomplete. We can say, however, that this necessarily less precise method of comparing estimates and marks is more satisfactory than basing an award on the result of a single external examination set to pupils in a small school or department of a school.

Extract from Conversion Table

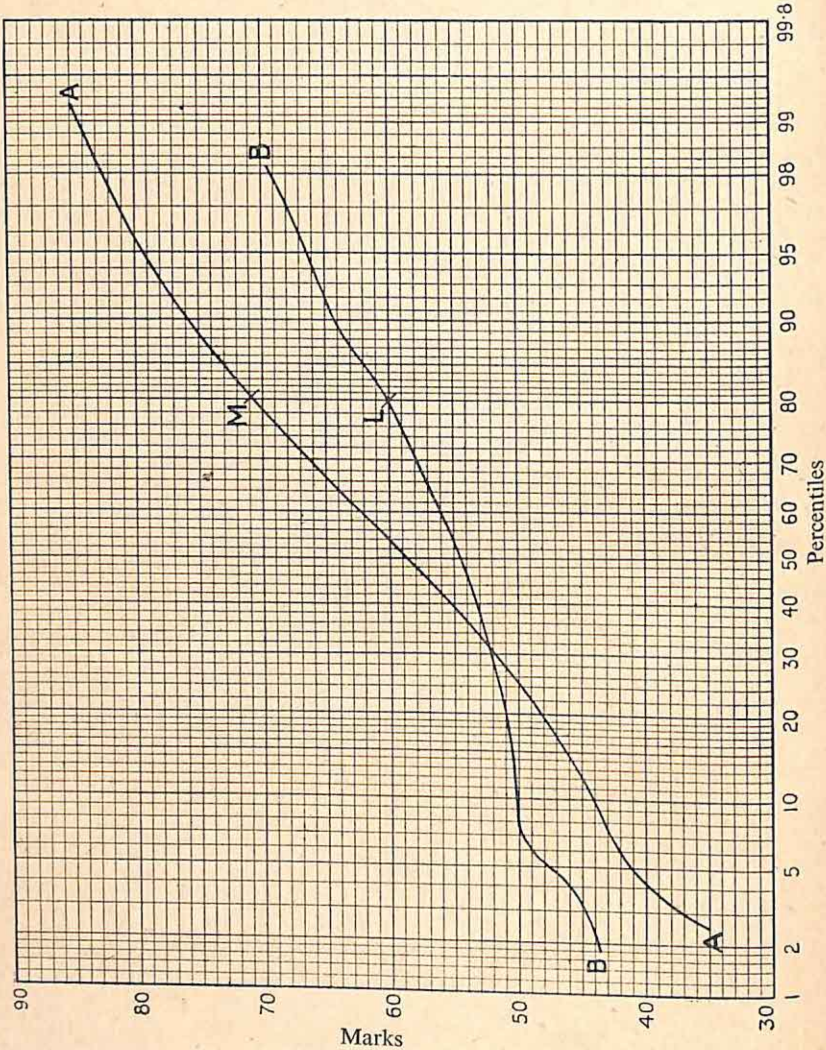
Estimates in English	
Raw	Scaled
50	43
51	48
52	52
53	55
54	58
55	60
56	62
57	65

Fundamental Principles

At the risk of wearying the reader we would draw his attention to the following fundamental principles which have been demonstrated in the above example:—

FIG. XX

Graphs for Scaling Leaving Certificate Teachers' Estimates



In Chapter XI we show how the scaling technique can be used with groups down to ten pupils.

Where the number of cases is less than ten, the teacher's standard of marking can be assessed to some extent by a careful study of the estimates, the test scores and the data on which the teachers have based their estimates. Our experience has shown that some basis for adjusting marks is possible when only a few candidates are under consideration. The relation between estimates and marks in the external test is often apparent when the corresponding estimates and marks are plotted on graph paper. Even with four or five marks, the examiner who has acquired an intimate

Extract from Conversion Table

Estimates in English	
Raw	Scaled
50	43
51	48
52	52
53	55
54	58
55	60
56	62
57	65

knowledge of how the scaling operation has worked over a large number of schools, can often give a good approximation to the alteration which must be made on the estimates. We do not propose to discuss this point in detail as our researches are still incomplete. We can say, however, that this necessarily less precise method of comparing estimates and marks is more satisfactory than basing an award on the result of a single external examination set to pupils in a small school or department of a school.

Fundamental Principles

At the risk of wearying the reader we would draw his attention to the following fundamental principles which have been demonstrated in the above example:—

(a) the teacher's order of merit is not altered by scaling.

(b) the pupil's score in the test is used only in the construction of a percentile graph; it does not determine whether the pupil passes or fails.

(c) the teachers' estimates in any subject are scaled on the results of a test or examination in the *same* subject.

(d) the scaling operation must be carried out school by school.

Teachers' Estimates

We should like to draw attention to Graph B in Figure XX. It will be seen that it "flattens out" after the mark 50 has been reached. This means that there are many pupils with marks just at or above the 50 level. At this critical 50 point the distribution of marks of 41 out of a group of 104 pupils was as follows:—

Mark	Frequency
53	10
52	9
51	11
50	9
49	1
48	0
47	1

Graph A, on the other hand, keeps the same slope all the way down. This is a persistent characteristic of Leaving Certificate estimates which makes them awkward to scale. Teachers tend to " give a pupil his chance." If he is lucky on the day of the examination he may score above the pass mark in the external examination and be credited with a pass. In addition, of course, most of the certain failures have been excluded from the examination. Scaling will be more accurate if teachers spread their marks at the lower end of the scale and if all pupils at this stage are included.

At the upper percentiles, Graph A is above Graph B; the pupils score higher marks than estimated by the teachers. This is another common feature of teachers' estimates at the

Leaving Certificate stage. Teachers tend to err on the safe side and undermark the abler pupils.

If the scaling procedure were to be adopted, teachers would require to give greater care to the distribution of their estimates. They would have, however, the satisfaction of knowing that greater attention would be paid to these estimates and that no freak result of an external examination would decide the fate of one of their charges.

CHAPTER X

CLASSROOM SCALING—ASSIGNING CLASS MEAN MARKS

A PARALLEL to the following incident must be within the knowledge of many teachers in secondary schools where it is the habit to allow unadjusted marks to go out to the public. Two boys, equally presentable in manners and appearance, applied for a job in a drawing office. One was a boy who was on the point of completing his three years' technical course (class mean IQ 110) and who had marks on his record card for the third year in the creditable 55-59 region. The other was a "C" group (class mean IQ 90) boy of small natural ability but very hard working. Under the pressure of strong parental encouragement he did well among his intellectual equals. The tests set to his class, based on their own schemes, were relatively easy and the marks comparatively high; his were among the highest. The conventional high marks for "C" pupils and moderate marks for "A" pupils were well understood within the school; both were considered to act as incentives. At the drawing office, naturally, both sets of marks were taken at their face value and the "C" boy was given the job. The result was inevitable: in a short time, all concerned, parent, employer and boy, found that the work was beyond his powers.

Clearly the time is ripe for exercising some control over the marks to be recorded for "C" pupils so as to bring them into some relationship to the marks recorded for pupils of known higher ability.

It would be educational folly to suggest as a method of control the setting of a common examination to all classes in the same year of study, a practice which can be adopted in highly selective senior secondary schools. In a school with pupils drawn from a wide range of intellectual ability where classes are following courses suited to the "age, ability and

aptitude" of the children, the examination for the "A" pupils would be quite beyond the capabilities of the "C" pupils. Some system of adjusting or scaling the examination marks within the school must be adopted.

Again, a teacher of a boys' section may have class marks in science with a mean of 65 while the teacher of the girls' section, which has the same general ability level, may have a class mean of 51. Each teacher may stoutly claim that his pupils have "earned" their marks. Some adjustment of the marks must be made in this case also if the marks of the two sections are to be made comparable.

The question whether these adjusted marks should be used only in the school records or on the pupils' record cards is one for the philosopher and not the scientist.

Within a school, class marks in the various subjects are made comparable by scaling. Comparability is achieved, as before, by making the mean and the scatter of the marks of each subject equal to some assigned mean and scatter. Two assumptions are made. Firstly, that class ability in any one subject is equal to that in any other; for example, class ability in English is assumed to be the same as the class ability in mathematics. Secondly, the variation of ability in one subject is equal to that in any other; for example, the range of ability in English is assumed to be the same as the range of ability in science. These assumptions agree with the common sense view when the numbers are large; it may be only approximately true for the usual school class of 30 to 40 pupils. Nevertheless, these assumptions are fundamental in the application of the scaling process to class marks.

Scaling school marks is a two-fold problem. In the first place we must scale the marks of each class in the various subjects; the possible variation in standard and scatter of the raw marks for a particular class was illustrated in Chapter II, page 11. Before we can scale subject marks we must determine what the mean mark for a particular class is to be and also what the standard deviation is to be.

We have assumed that the scatter of marks in each subject for all classes is the same and so our second problem becomes

the fixing of a mean mark for each class. If in a school there are six first year classes, 1A, IB, 1C, 1D, 1E and 1F, with differentiated schemes of work and examinations to suit the various ability levels, which class mean mark is to be 50? If class 1D has a class mean of 50 for all subjects, what will be the class mean for the other classes?

This chapter deals with a method for assigning the class means and the standard deviations of the class marks. The next chapter demonstrates how the scaling is done in practice. We denote the class mean mark by the symbol A.M.M. (assigned mean mark).

A good system of assigning means should have certain characteristics which we shall deal with in turn. Firstly, it would link the class mean mark in all subjects to some criterion indicating the general "brightness" of the class. Two criteria suggest themselves, (a) the mean class IQ and (b) the mean achievement mark in a promotion test. The mean class IQ is usually chosen as the more convenient index of brightness because IQs are on the same scale from year to year while actual achievement marks vary with the structure of the tests. There is, of course, a high degree of correlation between these two sets of marks.

Secondly, a good system of assigning class mean marks would fix a "key mark" which would have some educational significance. For example, on the Scottish National Pupil's Record Card there is an instruction that the standard of attainment reached in each subject will be represented by a percentage mark, "50 being taken to represent attainment that is only just satisfactory." Obviously the question immediately arises, "What type of pupil do the framers of this card have in mind?" Satisfactory attainment for a pupil in the senior secondary school is entirely different from satisfactory attainment in the junior secondary school. Again, satisfactory attainment in the "A" classes of the junior secondary school is not the same as in the "C" classes.

Investigation has shown that a pupil of IQ 110 has a fifty-fifty chance of gaining the Senior Leaving Certificate. For this type of school a class with a mean IQ of 110 might

be given a mean class mark of 50. If such a school had classes with a mean IQ of 100, the marks in a class of this nature would be depressingly low and would be regarded as "unsatisfactory."

McClelland's investigation showed also that a pupil with an IQ of 109 had a fifty-fifty chance of obtaining a Day School Certificate (Higher). If the mark of 50 were taken as the class mean mark for a class with mean IQ 109, the big majority of pupils in a Junior Secondary school fall into the category "unsatisfactory."

The third certificate mentioned in the McClelland investigation was the Day School Certificate (Lower) where the mark 50 was intended to mean "the just satisfactory completion of a two years' approved course." It was found that a pupil with an IQ of 94 had a fifty-fifty chance of achieving success in this certificate. This gives a good basis on which to build a scheme for assigning class means as it was reckoned that about two-thirds of the pupils in a junior secondary school normally gained this certificate. We shall take as the educational significance of a mark of 50 "the just satisfactory completion of a three years'* course" and assume that two-thirds of the pupils attain this standard.

The mean IQ of a large group of pupils selected at random is generally assumed to be 100 and the standard deviation 15. What shall we take as the class mean mark of a class with mean IQ 100? Reference to Table 20 shows that the sigma mark which divides a normal frequency distribution approximately into two-thirds and one-third is at roughly $0 \cdot 5\sigma$, from the mean. Hence if we make our class mean mark for a class with mean IQ 100 at $50 + 0 \cdot 5\sigma$, we have found what we are searching for.

Experience has shown that when all subjects—practical and academic—are taken into account, the average standard deviation of homogeneous classes is 10 and, if that value is accepted, our class mean mark corresponding to a class mean IQ of 100 will be $50 + \frac{1}{2} \times 10 = 55$.

* The school leaving age has been raised to 15 years since the establishment of this certificate.

Having fixed the educational significance of the mark of 50 and having chosen 10 as a suitable standard deviation, we must now find a formula for determining the assigned mean mark of any class. We shall assume that the interval between means of classes will be similar to the interval between their respective mean IQs. These intervals will not be equal because the measure of scatter of marks (10) is not equal to the measure for IQs (15) but these intervals should become equal if we express them in terms of their respective scatters or standard deviations. Since we know our key values—mean IQ of 100 corresponding to an assigned mean mark of 55—

$$\therefore \frac{\text{A.M.M.} - 55}{10} = \frac{\text{M.IQ} - 100}{15}$$

This gives us the scaling equation—

$$\text{A.M.M.} = 55 + \frac{10}{15}(\text{M.IQ} - 100)$$

For example, if a class mean IQ is 112, then

$$\begin{aligned} \text{A.M.M.} &= 55 + \frac{10}{15}(112 - 100) \\ &= 55 + 8 \\ &= 63 \end{aligned}$$

It ought to be emphasised that this formula, notwithstanding the above attempt to rationalise it, is essentially empirical; it crystallised out of an investigation designed to test whether a scale of marks based on an assigned mean mark of 55 for mean IQ 100 and 65 for mean IQ 115 would give marks in general agreement with current practice in a junior secondary school. It was found to work satisfactorily but we make no claim for this formula in other types of schools where the spread of ability is narrower.

When the second criterion—the mean promotion test score—is used, an almost identical value for the assigned mean mark is obtained. The following data refer to recent entrants to a junior secondary school. The standard deviation for the County in the promotion test was 50.

Class	Ia	Ib	Ic
Mean IQ Calculated assigned mean mark	113 64	100 55	91 49
Mean test mark Assigned mean mark	151 64	107 55	72 48

Calculation for class Ia—

$$\begin{aligned}\text{A.M.M.} &= 55 + \frac{10}{50}(151 - 107) \\ &= 55 + 8{\cdot}8 \\ &= 64\end{aligned}$$

We can also find out what percentage of the class is likely to fall below 50.

$$\begin{aligned}\text{Distance of 50 below class mean} &= 63 - 50 \\ &= 13 \\ &= \frac{13}{10}\sigma \\ &= 1{\cdot}3\sigma\end{aligned}$$

From Appendix IV, $-1{\cdot}30\sigma$ corresponds to the 9·7 percentile, *i.e.*, 9·7 per cent. would be below 50.

The following gives data for typical classes:—

Group	Mean IQ	Assigned Mean Mark	Percentage below 50
Typical A	112	63	10
,, B	100	55	31
,, C	90	48	58

A third characteristic of a good system of assigning class means would be that the overlap between the marks given to typical A, B and C classes would not be too great. A system that allowed for too great an overlap would not attain its purpose and the drawing office episode might occur again. On the other hand, too small an overlap might make it appear as if A, B and C were different species instead of belonging to one continuous variation. The following warning from

Thorndike* makes this point:—" It should be unnecessary to warn the reader against the absurdity of deliberately changing continuous variation into a few groups by coarse scaling; next assuming that the central part of one of these coarse divisions really measures all the individuals therein, and finally imagining that because the continuous series varying from "a" to "a + b" has been called say, poor, medium, good and excellent, there are really gaps within it! Unfortunately even gifted thinkers are guilty of this error."

TABLE 13

Typical Distributions of Scaled Marks

Group	A	B	C	Total
No. of classes	7	6	4	17
Average assigned mean mark	64	55	48	57
Class interval	Frequency			
90-94	1			1
85-89	6			6
80-84	6	2		8
75-79	20	3		23
70-74	29	7		36
65-69	28	21	4	53
60-64	29	30	15	74
55-59	41	34	13	88
50-54	23	35	20	78
45-49	10	29	22	61
40-44	5	11	19	35
35-39	1	8	11	20
30-34		1	6	7
25-29			3	3
20-24		1		1
Total	199	182	113	494
Below 50	8%	27%	52%	26%

It would be absurd, indeed, if the best "C" boy at his special aptitude subject were, by the incidence of a scaling procedure, prevented from rising as high in marks as the median boy in a typical "A" class.

If we take the three classes indicated above as being representative of their groups, in the normal case there will be

* THORNDIKE, *Educational Psychology*, page 150, (Kegan Paul).

considerable overlap in the marks of the three classes. Twenty per cent. of the Bs will reach or exceed the mean of the As and 24 per cent. of the Cs will exceed the mean of the Bs.

Table 13 shows typical distributions of scaled marks in A, B and C group classes. A study of this table shows the extent of the overlap of the marks between the three groups.

The method of assigning class mean marks outlined above has proved to be successful in practice. No doubt, once the system of scaling has been widely adopted, some modifications will suggest themselves. If this formula or any other which is found to be more feasible were adopted in junior secondary schools, marks would be comparable from school to school.

CHAPTER XI

SCALING IN THE CLASSROOM

THE scaling procedure in the classroom involves transforming the teacher's raw examination marks into marks with an assigned mean and standard deviation. We have indicated in the previous chapter how to arrive at the assigned mean and that the assigned standard deviation for all subjects is taken as 10.

The scaling equation for this transformation would be

$$X_S = \text{A.M.M.} + \frac{10}{\sigma_X}(X - M)$$

where X_S = scaled examination mark

σ_X = standard deviation of class marks

X = class mark

M = mean of class marks

$(X - M)$ is the deviation of a pupil's mark from the mean of the class marks; for example, where the class mean mark in arithmetic is 50, the pupil with a mark of 65 has a deviation of 15 from the mean. In the above equation $\frac{10}{\sigma_X}$ may be regarded as the adjusting ratio applied to the pupil's deviation from the mean. To find the scaled examination mark the adjusted deviation from the mean is added to the assigned mean mark.

An adjusting Table (Appendix VII) has been compiled to save the teacher's time and gives adjusted deviations for all cases that are likely to arise in the class scaling procedure.

So far we have assumed that the mean and standard deviation have been calculated in the usual way. We feel, however, that such calculations may be considered too lengthy and that a suitable estimate of their values will have to be found to save teachers' time even if some sacrifice of precision has to be made. Fortunately, as Garrett has pointed out, the virtual equality of theoretically more accurate and

theoretically less accurate methods is a familiar finding in mental measurements.*

In Chapter III we indicated that in a normal distribution P_{16} and P_{84} cut off 16% of the cases from either end of the distribution; these percentiles are at a distance -1σ and $+1\sigma$ respectively from the mean. The range $P_{84} - P_{16}$ (to be named E) may be taken as an estimate of 2σ and both theoretical considerations and investigations of actual class marks have shown that even with distributions not entirely normal, this range is a sufficiently good estimate of 2σ for scaling purposes.

There are three methods of classroom scaling, the first two based on different ways of calculating $P_{84} - P_{16}$.

METHOD 1—suitable for the case of whole classes of 25 pupils and over and deemed to be passably normal.

If the pupils' papers are ranked in order of marks from the highest, the class rank corresponding to any percentile rank may be obtained in the following way:—

$$\text{Class rank} = \frac{x}{100} \times N + 0{\cdot}5$$

where x is the percentile and N the number of pupils. For example, the class rank corresponding to P_{16} in a class of 26 pupils is $\frac{16}{100} \times 26 + 0{\cdot}5$ which is 4·7. Accordingly P_{16} lies about half-way between the scores of the fourth and fifth pupils from the bottom. To get the smoothed value—the value that will reflect the general trend of the marks in this region—we take the average of the scores of the 3rd, 4th, 5th and 6th pupils from the bottom. Similarly, to calculate P_{84} we take the average of the 3rd, 4th, 5th and 6th pupils from the top, and so with other cases. To save the teacher the trouble of calculation, Appendix VIII has been compiled giving class ranks for certain useful percentiles.

The scaling procedure by Method 1 will be illustrated by transforming the 27 marks obtained in a history examination by a class whose assigned mean mark is 63.

* GARRETT. *Statistics in Psychology and Education*, 2nd Edition, 1944, p 166. 3rd Edition, 1947, p 169, (Longman Green and Co.).

TABLE 14

History Examination Marks for 27 Pupils

Order	Raw Marks	Order	Raw Marks	Order	Raw Marks
1	85	11	56	21	46
2	81	12	54	22	46
3	78	13	53	23	40
4	73	14	53	24	40
5	65	15	52	25	34
6	65	16	52	26	33
7	63	17	51	27	31
8	60	18	50		
9	59	19	47		
10	59	20	47		

Step 1—Arrange pupils' papers in order of merit, the highest scoring pupil being ranked 1.

Step 2—Ascertain the assigned mean mark for the class according to the method outlined in Chapter X. In this example we are told that the A.M.M. is 63.

Step 3—From Appendix VIII the class rank corresponding to P_{84}/P_{16} is 4·8. Therefore the percentile scores lie between the fourth and fifth scores from the top and bottom.

Step 4—Find the average of four marks, two on either side of P_{84}. This mark may be taken as the mark of the 84th percentile pupil. Count off the 3rd, 4th, 5th and 6th from the top, 78, 73, 65, 65. $P_{84} = 70$, the average of those four marks.

Step 5—Find P_{16} in the same way. $P_{16} = 40$, the average of 34, 40, 40 and 46.

Step 6—Calculate the mean of the marks by taking the average of P_{84} and P_{16}.

$$\text{Mean} = \frac{P_{84} + P_{16}}{2} = \frac{70 + 40}{2} = 55$$

Step 7—Calculate the standard deviation $\sigma = \frac{E}{2} = \frac{P_{84} - P_{16}}{2}$

$$\therefore \sigma = \frac{70 - 40}{2}$$
$$= 15$$

Step 8—Find the adjusting ratio—$\dfrac{\text{assigned sigma}}{\text{marks sigma}}$

$$\text{adjusting ratio } = \frac{10}{15} = \frac{2}{3}$$

Step 9—Scale each raw mark in turn by finding its deviation from the mean, applying the adjusting ratio and adding to the A.M.M.

For example, 85 is $+30$ from the mean 55. Applying the adjusting ratio $30 \times \frac{2}{3} = 20$. Hence the scaled mark becomes $20 + \text{A.M.M.}$ which is $20 + 63 = 83$.

Similarly with the mark 81—

$$\text{Scaled mark} = 63 + \frac{2}{3} \times 26 = 63 + 17 = 80$$

Practical Aids

1. Appendix VII gives the application of the adjusting ratio to the deviation from the mean for a wide range of

DATA SHEET I

No. in class............... A.M.M..................................

From Table 25 rank for P84/P16 = ...

P84 scores (1)
(2)
(3)
(4) ________

Sum ________

P16 scores (1)
(2)
(3)
(4) ________

Sum ________

Mean P84....................... Mean P16....................................

P84–P16 (E) = ...

Adjusting ratio $= \dfrac{10}{E/2} =$

Mean score $= \dfrac{P84 + P16}{2} =$

ratios and deviations. Reference to this table will save the teachers' time.

2. It is advisable to keep a record of the calculations for each set of papers. This may be done by using Data Sheet I (see page 83) which should be attached to the examination papers.

3. On each examination paper the following rubber stamp imprint is made:—

FIG. XXI

Rubber Stamp Imprint for Scaling Calculation

Raw mark	
Mean	
Deviation	
Adjusted Dev.	
A.M.M.	
Scaled mark	

Once Data Sheet I has been completed it is a relatively simple job with the help of Appendix VII to complete the entries on each examination paper.

The reader is invited to work out the scaled marks in the above example to give him an estimate of the time taken to scale the marks of a particular case. The scaled marks are:—

83, 80, 78, 75, 70, 70, 68, 66, 65, 65, 63, 62, 61,
61, 61, 61, 60, 60, 57, 57, 57, 57, 53, 53, 49, 48, 47.

METHOD 2—suitable for half classes from 12 to 24 pupils whose marks are assumed to be distributed in passably normal fashion. Investigation has shown that this method works well with "judgment" marks in domestic subjects, art and handcrafts.

The calculation of P_{84} and P_{16} depends on what is known as the centroid property of the normal curve. It is known that the mean of the highest 38% of the scores in a normal distribution has a deviation of σ from the mean and

thus to find P_{84} we take the average of the top 38% of the marks. For example, in a class of 22, P_{84} is obtained by finding the average of the top eight marks. Similarly P_{16} is given by the average of the bottom eight marks. Appendix IX gives the number of marks that have to be averaged from top and bottom to give an estimate of P_{84} and P_{16}.

The scaling procedure by Method 2 is illustrated by the following example:—

TABLE 15

Benchwork Marks of a Class of 15 *Boys*

Order	Mark	Order	Mark
1	76	9	61
2	70	10	58
3	69	11	55
4	68	12	54
5	67	13	52
6	66	14	51
7	64	15	49
8	61		

Step 1—Arrange examination papers in order of merit, the highest scoring pupil being ranked 1.

Step 2—Ascertain the A.M.M. by the method outlined in Chapter X. In this case the A.M.M. is 58.

Step 3—Estimate P_{84} by averaging the required number of marks from the top. The number of marks to be averaged is given in Appendix IX.

From this table we note that six marks have to be averaged.

$$\therefore P_{84} = \frac{1}{6}(76 + 70 + 69 + 68 + 67 + 66)$$
$$= 69$$

Step 4—Estimate P_{16}.

$$P_{16} = \frac{1}{6}(49 + 51 + 52 + 54 + 55 + 58)$$
$$= 53$$

Step 5—Calculate standard deviation.

$$\frac{E}{2} = \sigma = \frac{P_{84} - P_{16}}{2}$$

$$\therefore \sigma = \frac{69 - 53}{2} = 8$$

Step 6—Calculate the mean mark.

$$\text{Mean} = \frac{P_{84} + P_{16}}{2}$$

$$\text{Mean} = \frac{69 + 53}{2} = 61$$

Step 7—Find adjusting ratio—$\dfrac{\text{assigned } \sigma}{\text{marks } \sigma}$.

$$\text{Adjusting ratio} = \frac{10}{8}$$

Step 8—Scale each raw mark by finding its deviation from the mean, applying the adjusting ratio and adding to the A.M.M.

Mark 76 is 15 above the mean

$$\therefore \text{scaled mark} = 58 + \frac{10}{8} \times 15$$
$$= 77$$

Practical Aids

1. Use Data Sheet II for calculations. Data Sheet II is shown completed for the above example.

Data Sheet II

No. in class—15. A.M.M.—58.

From Table 26 number of marks to be averaged for P84 and P16—6.

P84 marks		P16 marks	
1.	76	1.	49
2.	70	2.	51
3.	69	3.	52
4.	68	4.	54
5.	67	5.	55
6.	66	6.	58
7.		7.	
8.		8.	
9.		9.	
Sum	416	Sum	319
Average	69	Average	53

$$\sigma = \frac{E}{2} = \frac{P84 - P16}{2} = \frac{16}{2} = 8$$

$$\text{Mean mark} = \frac{P84 + P16}{2} = \frac{122}{2} = 61$$

$$\text{Adjusting ratio} = \frac{10}{8}$$

2. Print rubber stamp outline of Figure XXI on each examination paper.

3. Table 24 gives the application of the adjusting ratio to the deviation from the mean.

METHOD 3—suitable for cases where the distribution of marks is not thought to be normal, specially for half class subjects except those mentioned under Method 2.

This method, which is a graphical one, is also based on the centroid property of the normal distribution and Method 2 may be considered as a special case of this method. It has also the advantage that it makes use of all marks.

The Data Sheet for Method 3 is as follows:—

DATA SHEET III

No. in class.......................... A.M.M.................

From Table 27.

		( ) A	( ) B	( ) C	( ) D	( ) E
Scores						
	1					
	2					
	3					
	4					
	5					
Sums						
Means						
Sigma Deviations						

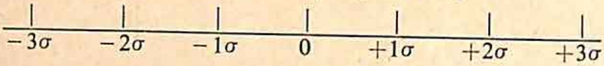
The completion of the above data sheet is best illustrated by an example. Take the case of 18 examination marks. These may be divided into 5 groups, 3, 4, 4, 4, 3, corresponding to percentages 17, 22, 22, 22, 17. If the groups are named A, B, C, D and E their means are at the following sigma distances from the mean 0.

(3)	(4)	(4)	(4)	(3)
A	B	C	D	E
17%	22%	22%	22%	17%
$-1{\cdot}49\sigma$	$-0{\cdot}59\sigma$	0	$+0{\cdot}59\sigma$	$+1{\cdot}49\sigma$

The graphical method employed is not the normal one but is known as the sigma chart. The vertical axis is marked off in the usual way but the horizontal axis has the zero in the centre and the units to the right and left are sigma units as in Figure XXII below.

FIG. XXII

Horizontal Axis of Sigma Chart



The various steps in Method 3 are illustrated by the following example :—

Scale the 14 class marks, for which 61 is the A.M.M. :—

33, 47, 48, 52, 52, 52, 53, 54, 55, 56, 56, 63, 67, 76.

Step 1—Arrange pupils' papers in order or marks, the *lowest* mark on top.

Step 2—Ascertain the assigned mean mark as shown in Chapter X. In the above case A.M.M. is 61.

Step 3—From Table 27 find the grouping of the marks and the sigma deviation of each group. For a class of 14 this is—

A	B	C	D	E
3	4	4	3	0
−1·37	−0·38	+0·38	+1·37	

DATA SHEET III

No. in class—14. A.M.M.—61

From Table 27.

Scores	(3) A	(4) B	(4) C	(3) D	(0) E
1	33	52	54	63	
2	47	52	55	67	
3	48	52	56	76	
4		53	56		
5					
Sums	128	209	221	206	
Means	43	52	55	69	
Sigma Devs.	−1.37	−0.38	+0.38	+1.37	

Step 4—Complete Data Sheet III.

Note that in this example there are no entries under E.

Step 5—On a sheet of graph paper lay off a sigma scale along the horizontal axis

$$-2\sigma,\ -1\sigma, 0,\ +1\sigma,\ +2\sigma\ (1 \text{ sigma} = 2 \text{ ins.})$$

Step 6—Along the vertical axis lay off a scale of examination marks. Repeat this scale at 0 in the sigma scale. Scale vertically 10 marks to the inch. (Figure XXIII).

Step 7—Plot A.M.M. above 0 and (A.M.M. + 10) above $+1\sigma$ and draw a straight line through them. This is the line of scaled marks. In Figure XXIII the points are P and Q where P is 61 above 0 and Q is 71 above $+1\sigma$.

Step 8—Plot A, B, C, D and E above the sigma deviations. A is 43 above $-1{\cdot}37\sigma$, *i.e.*, $(-1{\cdot}37,\ 43)$, B $(-0{\cdot}38,\ 52)$, C $(+0{\cdot}38,\ 55)$, D $(+1{\cdot}37,\ 69)$. Join these points by straight lines to form the line of marks.

Step 9—Scale the raw marks by reading from the line of raw marks to the line of scaled marks.

In Figure XXIII the line of arrowheads shows how the raw mark of 53 when scaled becomes 60.

The following Table gives the results of scaling:—

TABLE 16

Raw and Scaled Marks for a Class of 14 Pupils

Raw Marks	Scaled Marks	Raw Marks	Scaled Marks
33	37	54	62
47	52	55	65
48	53	56	65
52	57	56	65
52	57	63	71
52	57	67	73
53	60	76	80

GRAPHICAL TECHNIQUE FOR METHODS 1 and 2.

Teachers find the rubber stamp technique quick and easy but those who are familiar with graphs prefer this method. We shall illustrate this approach by the same example which was used in Method 1.

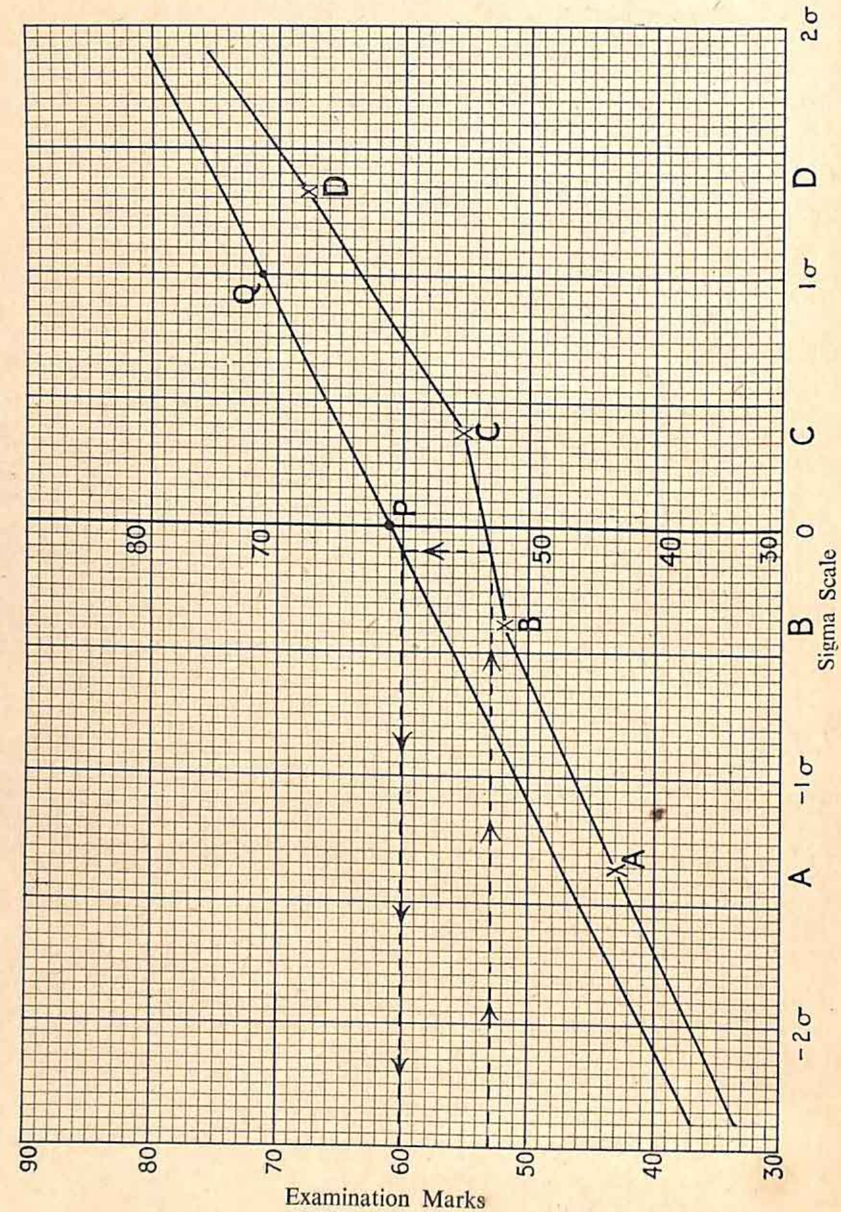


FIG. XXIII
Sigma Chart for Scaling 14 Class Marks

1. We assume that A.M.M., P_{84} and P_{16} have been calculated by the appropriate technique. In the example given :

$$\text{A.M.M.} = 63$$
$$P_{84} = 70$$
$$P_{16} = 40$$

2. Lay off a sigma scale along the horizontal axis, 1 sigma = 2 inch. (Figure XXIV).

3. Let the vertical axis represent examination marks. This scale should be repeated at the zero of the sigma scale. In Figure XXIV the vertical scale is 10 marks to the inch.

4. Plot P_{16} vertically above -1σ and P_{84} above $+1\sigma$ and draw a line through these two points—the line of marks. P_{16} is the point $(-1\sigma, 40)$ and P_{84} is $(+1\sigma, 70)$.

5. Plot the A.M.M. above 0 and (A.M.M. + 10) above $+1\sigma$ and draw a line through them—the line of scaled marks. In Figure XXIV the points are P (0, 63) and Q $(+1\sigma, 73)$.

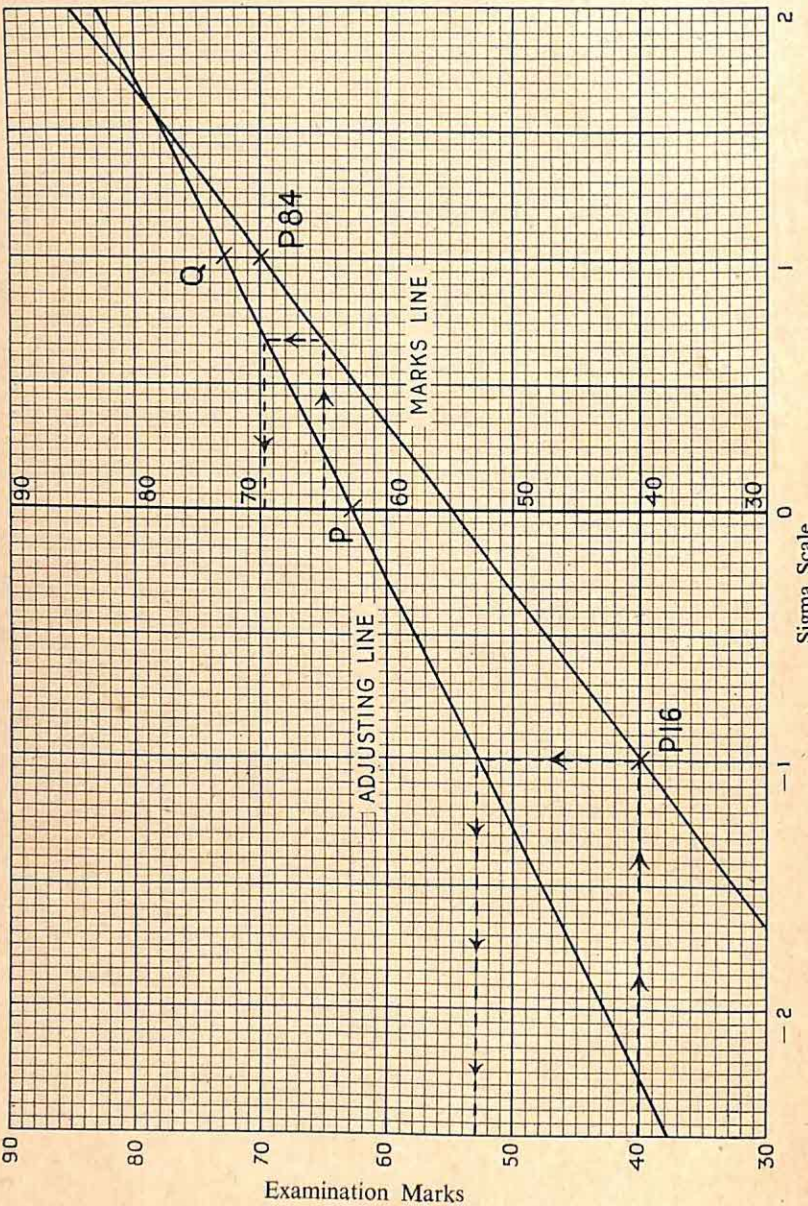
6. Read from the line of raw marks to the line of scaled marks. In Figure XXIV we have illustrated the scaling of the raw mark 65 which becomes a scaled mark of 70 and also 40 which scales to 53.

Class Averages

Paragraph 699 of the Advisory Report on Secondary Education refers in the last sentence to what it calls "the usual invalid method" of totalling and averaging marks to find a pupil's general place in class. Scaled marks, of course, can be validly added; that is one of the purposes of scaling. A difficulty, however, arises when marks—scaled or unscaled—are totalled and averaged: the scatter of the marks shrinks considerably, the shrinkage increasing with the number of marks added. A way out is to apply Method 1 in order to restore the scatter to its original standard deviation of 10. Unless this is done an average of 50 is no longer "the lower limit of satisfactory performance" although the average was derived from marks in which 50 had that significance.

FIG. XXIV

Graphical Method for Scaling History Marks for 27 Pupils



POSTSCRIPT

We hope that many practising teachers and administrators will have stayed the journey thus far. Their reward for so doing will be a greater knowledge of the significance of examination marks but we would strongly assert that this knowledge can only be enriched if they put the scaling operation into practice.

For practical reasons, examinations will continue to have their place. *Inside* the classroom examination marks are the normal means of communication between the teacher and his class and accordingly need not be scaled. They must be scaled when they have to be compounded, compared and interpreted in places *outside* the classroom, in the headmaster's room, in the home, in the vocational guidance department, in the offices of prospective employers and in other schools. Furthermore, now that the great majority of pupils leave school without any externally accredited certificate, it is of first importance that the mark on the leaver's card should be based on a planned system of marking: that means scaled marks. We are convinced and we hope that those who have borne with us thus far are also convinced that the scaling process is essential to make many examination marks have a meaning.

Perhaps readers will pardon us if we re-state a part of our educational philosophy which has earlier been written in this book. Examinations form only a part of the educative process and not the most important part. Good taste, character and citizenship must all be cultivated in the schools and these do not submit to measurement. Again, despite advances in psychology, the human personality defies and in our opinion, will always defy exact measurement. It is in the interest of humanity that this should be so.

Finally, we make no extravagant claims for our methods. There is still much to be investigated and we are sure that those who follow our suggestions will improve upon our

techniques. We are at present experimenting with methods applicable to small numbers of pupils, with short cuts and approximations designed to lighten, without too great a loss of accuracy, the admittedly heavy labour of some methods and with special investigations of individual cases where the procedure has produced unusual results.

This book was written at the request of many who had heard of our attempts at introducing scaling as a working system. We offer it as a contribution to the knowledge which experimental education has given to the teacher. In the last 30 years or so, this branch of education has enabled teaching to be more effective, schools to be happier and the aims of education to be more thoroughly realised. We know only too well that in solving one problem we have created others: to solve the problem of to-day is to raise the problem of to-morrow.

APPENDIX I

EXAMPLES OF OBJECTIVE TESTS IN ENGLISH, LATIN AND MATHEMATICS

These tests were applied to pupils of about 17 *years of age who had completed a five years secondary course.*

English

A. In each of the following sentences *one* word is wrongly employed. Underline that word and, in the space provided at the end of the sentence, write the word that should have been used.

(1) Late that night we saw a curious phenomena ....
(2) Each of the girls came out in their turn..........
(3) Neither men nor money were lacking for our plans..........
(4) He had seldom or ever seen such flowers..........
(5) You have weakened instead of strengthened your case..........
(6) Her coat is as fine than but less expensive than mine..........
(7) The interest of the visitors centred round the old keep..........
(8) "Whom do you think you are?" she retorted..........
(9) A good employer is one of those who never questions trusted employees ....
(10) Nothing but acting and games interest him..........

B. Complete each of the following sentences by inserting in the blank space what you consider the most suitable preposition.

(1) .... a pinch we could do without tobacco for good.
(2) The solution of so difficult a problem was quite .... him.
(3) To pay the bill he drew a cheque .... his bank.
(4) Her geometry was good but she was weak .... algebra.
(5) He was an amiable man, slow .... anger.
(6) Political speeches are mainly .... the heads of children.
(7) Although he was innocent, his name was .... a cloud.
(8) .... my part, I shall give you all the help I can.
(9) It is difficult to get a white elephant .... one's hands.
(10) He is .... nature a taciturn man.

C. Strength; condescension; cowardice; humour; weather; frankness; friendship; endurance; darkness; revels; knowledge; cunning; stride; honesty; furnishings.
From the above list of nouns choose that which will most suitably pair with each of the following adjectives. Write the chosen noun in the space provided after the adjective.

1. Stygian ....
2. Spartan ....
3. Machiavellian ....
4. Halcyon ....
5. Martial ....
6. Olympian ....
7. Jacobean ....
8. Platonic ....
9. Bacchanalian ....
10. Rabelaisian ....

LATIN

Read this passage, mentally translating as you read, then answer the questions which follow:—

Multa super Lauso regitat, multumque remittit
qui revocent maestique ferant mandata parentis.
at Lausum socii exanimem super arma ferebant
flentes, ingentem atque ingenti vulnere victum.
agnovit longe genitum praesaga mali mens.
canitiem multo deformat pulvere et ambas
ad caelum tendit palmas et corpore inhaeret.
'tantane me tenuit vivendi, nate, voluptas,
ut pro me hostili paterer succedere dextrae,
quem genui? tuane haec genitor per vulnera servor
morte tua vivens? heu, nunc miser mihi demum
exitium infelix, nunc alte vulnus adactum!

(a) State and explain the mood of "revocent" (verse 2).

....

(b) State and explain the case of "mali" (verse 5).

....

(c) His comrades brought Lausus home—exhausted, breathless, distraught, dead, wounded. Underline *one* of these words as your answer.

(d) Of what does Lausus' father accuse himself?

....

(e) What attitude of prayer is here described?

....

(f) Scan verse 4:—
flentes, ingentem atque ingenti vulnere victum.

(g) What is the first action of the father, on learning what has happened?

....

(h) Who is meant by "quem genui"?

....

(i) What are you told about the personal appearance of the old man?

....

(j) What first led him to guess what had happened?

....

MATHEMATICS

Attempt as many questions as possible. If you require to do any working, use the blank spaces provided for this purpose.

Answers

1. (a) Find in square inches the area of a circle of radius 7 inches. (Take $\pi=\frac{22}{7}$) — Area=
   (b) Find, in c.c., the volume of a cone with base 6 sq. cm. and height 8 cm. — Vol.=

2. (a) Express $\sqrt{\text{Cosec}^2A-1}$ as a trig. ratio of A. — (a)
   (b) If $\text{Sin}^2B \equiv (1-x)\ (1+x)$, express x as a trig. ratio of B. — x=
   (c) Give the values of :—
   (i) Sin 90° — Sin 90°=
   (ii) Cos 180° — Cos 180°=

3. Replace each of the following by a single trig. ratio:—
   (a) 2 Sin C. CosC. — (a)
   (b) $\frac{\tan x-\tan y}{1+\tan x \tan y}$ — (b)
   (c) $\frac{1-\tan^2\frac{\theta}{2}}{1+\tan^2\frac{\theta}{2}}$ — (c)
   (d) $\frac{b^2+c^2-a^2}{2bc}$ where a, b, c are sides of triangle ABC. — (d)

4. (a) Find the two solutions for x:— — 1st root=
   $ax^2+(a^2-1)\ x-a=0$ — 2nd root=
   (b) $\left.\begin{aligned} x^2+xy &= 12 \\ xy+y^2 &= 4 \end{aligned}\right\}$ Find the positive values of (x+y) and x. — x+y= ; x=

5. (a) Find the positive square root of $4+2\sqrt{3}$ (Answer in surd form). — (a)
   (b) Find, in its simplest form, the value of
   $\frac{1}{\sqrt{3}-1}-\frac{1}{\sqrt{3}+1}$ — (b)

APPENDIX II

TABLE 17

Calculation of the Mean from Grouped Data

Class Interval	f	d	fd
90-94	1	6	6
85-89	0	5	0
80-84	0	4	0
75-79	1	3	3
70-74	9	2	18
65-69	11	1	11
60-64	15	0	0
55-59	10	−1	−10
50-54	11	−2	−22
45-49	9	−3	−27
40-44	3	−4	−12
35-39	1	−5	−5
N	71		+38 −76
			Σfd − 38

$$c=5$$

$$A=62$$

$$\frac{\Sigma fd}{N}=\frac{-38}{71}$$

$$=-0.54$$

$$\text{Mean}= A+\frac{\Sigma fd}{N}\times c$$

$$=62-0.54\times 5$$

$$=62-2.7$$

$$=59.3$$

Step 1—Select an assumed or guessed mean, A, at the mid-point of the interval which is most likely to contain the actual mean. In the above example, A is at the mid-point of the class interval 60-64, *i.e.*, A=62.

Step 2—In the column headed "d" indicate the number of intervals between each interval mid-point and the assumed mean. Note that below the assumed mean the values are negative.

Step 3—Multiply each frequency "f" by its corresponding "d" to give the products in column "fd."

Step 4—Find the algebraic sum of the fd products to give Σfd.

Step 5—Divide Σfd by N, the total number of cases.

Step 6—Multiply Σfd by the size of the class interval, c.

Step 7—Add $\frac{\Sigma fd}{N}\times c$ algebraically to A to give the correct mean.

APPENDIX III

TABLE 18

Calculation of the Standard Deviation from Grouped Data

Class Interval	f	d	fd	fd^2
90-94	1	6	6	36
85-89	0	5	0	0
80-84	0	4	0	0
75-79	1	3	3	9
70-74	9	2	18	36
65-69	11	1	11	11
60-64	15	0		
55-59	10	−1	−10	10
50-54	11	−2	−22	44
45-49	9	−3	−27	81
40-44	3	−4	−12	48
35-39	1	−5	−5	25
N	71		+38 −76	291
			$\Sigma fd - 38$	

$$c = 5$$

$$A = 62$$

$$\frac{\Sigma fd}{N} = \frac{-38}{71} = -0.54$$

$$\left(\frac{\Sigma fd}{N}\right)^2 = (-0.54)^2 = 0.29$$

$$\frac{\Sigma fd^2}{N} = \frac{291}{71} = 4.10$$

$$\sigma = \left[\sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}\,\right] \times c$$

$$= \left[\sqrt{4.10 - 0.29}\,\right] \times c$$

$$= 1.95 \times 5$$

$$= 9.75$$

Step 1—Complete the d and fd columns as in the calculation of the mean.

Step 2—Multiply each fd by its corresponding d to give the products in the fd^2 column.

Step 3—Find the sum of the fd^2 products, denoted by Σfd^2.

Step 4—Find $\left(\frac{\Sigma fd}{N}\right)^2$ *i.e.* $\left(\frac{-38}{71}\right)^2$ and $\frac{\Sigma fd^2}{N}$ *i.e.* $\frac{291}{71}$

Step 5—The standard deviation is given by the formula—

$$\sigma = \left[\sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}\,\right] c$$

APPENDIX IV

TABLE 19

Percentiles for given Standard Deviations

S.D. Value	Percentiles	S.D. Value	Percentiles	S.D. Value	Percentiles
−3.3	.05	−1.1	13.6	1.1	86.4
−3.2	.07	−1.0	15.9	1.2	88.5
−3.1	.10	− .9	18.4	1.3	90.3
−3.0	.14	− .8	21.2	1.4	91.9
−2.9	.19	− .7	24.2	1.5	93.3
−2.8	.26	− .6	27.4	1.6	94.5
−2.7	.35	− .5	30.8	1.7	95.5
−2.6	.47	− .4	34.5	1.8	96.4
−2.5	.62	− .3	38.2	1.9	97.1
−2.4	.82	− .2	42.1	2.0	97.7
−2.3	1.1	− .1	46.0	2.1	98.2
−2.2	1.4	0	50.0	2.2	98.6
−2.1	1.8	.1	54.0	2.3	98.9
−2.0	2.3	.2	57.9	2.4	99.18
−1.9	2.9	.3	61.8	2.5	99.38
−1.8	3.6	.4	65.5	2.6	99.53
−1.7	4.5	.5	69.2	2.7	99.66
−1.6	5.5	.6	72.6	2.8	99.74
−1.5	6.7	.7	75.8	2.9	99.81
−1.4	8.1	.8	78.8	3.0	99.87
−1.3	9.7	.9	81.6	3.1	99.90
−1.2	11.5	1.0	84.1	3.2	99.93
				3.3	99.95

APPENDIX V

TABLE 20

Standard Deviations for given Percentiles

Percentiles	S.D. Value	Percentiles	S.D. Value	Percentiles	S.D. Value
.1	−3.090	31	−.496	70	+.524
.2	−2.878	32	−.468	71	+.553
.3	−2.748	33	−.440	72	+.583
.4	−2.652	34	−.412	73	+.613
.5	−2.576	35	−.385	74	+.643
.6	−2.512	36	−.358	75	+.675
.7	−2.457	37	−.332	76	+.706
.8	−2.409	38	−.305	77	+.739
.9	−2.366	39	−.279	78	+.772
1	−2.326	40	−.253	79	+.806
2	−2.054	41	−.228	80	+.842
3	−1.881	42	−.202	81	+.878
4	−1.751	43	−.176	82	+.915
5	−1.645	44	−.151	83	+.954
6	−1.555	45	−.126	84	+.995
7	−1.476	46	−.100	85	+1.036
8	−1.405	47	−.075	86	+1.080
9	−1.341	48	−.050	87	+1.126
10	−1.282	49	−.025	88	+1.175
11	−1.227	50	−.000	89	+1.227
12	−1.175	51	+.025	90	+1.282
13	−1.126	52	+.050	91	+1.341
14	−1.080	53	+.075	92	+1.405
15	−1.036	54	+.100	93	+1.476
16	−.995	55	+.126	94	+1.555
17	−.954	56	+.151	95	+1.645
18	−.915	57	+.176	96	+1.751
19	−.878	58	+.202	97	+1.881
20	−.842	59	+.228	98	+2.054
21	−.806	60	+.253	99	+2.326
22	−.772	61	+.279	99.1	+2.366
23	−.739	62	+.305	99.2	+2.409
24	−.706	63	+.332	99.3	+2.475
25	−.675	64	+.358	99.4	+2.512
26	−.643	65	+.385	99.5	+2.576
27	−.613	66	+.412	99.6	+2.652
28	−.583	67	+.440	99.7	+2.748
29	−.553	68	+.468	99.8	+2.878
30	−.524	69	+.496	99.9	+3.090

APPENDIX VI

ARITHMETICAL METHOD OF SCALING

(a) For use where the number of pupils is relatively small

THIS method which involves a direct application of the scaling equation (see page 36) is best illustrated by an example. In Table 21 below, column 2 shows the teacher's estimate in English for each of the 20 pupils in the class: column 4 shows the mark obtained by the pupil in the external examination on which the marks are to be scaled.

TABLE 21

Pupil	Estimate	Square	Mark	Square
1	40	1600	47	2209
2	45	2025	54	2916
3	65	4225	66	4356
4	50	2500	41	1681
5	60	3600	59	3481
6	65	4225	74	5476
7	80	6400	84	7056
8	35	1225	40	1600
9	42	1764	49	2401
10	45	2025	54	2916
11	48	2304	56	3136
12	70	4900	72	5184
13	45	2025	49	2401
14	45	2025	56	3136
15	60	3600	78	6084
16	40	1600	43	1849
17	60	3600	66	4356
18	35	1225	32	1024
19	40	1600	41	1681
20	60	3600	65	4225
20 (n)	1030 (E1)	56068 (E2)	1126 (S1)	67168 (S2)

The steps in the calculations are as follows:—

Step 1—Enter in column 3 the square of each estimate.

Step 2—Enter in column 5 the square of each mark.

Step 3—Add columns 2, 3, 4 and 5 to find the values of E1, E2, S1 and S2. The number of pupils (n) is given at the foot of column 1.

Step 4—Calculate the average estimate, *i.e.*, $\frac{\text{E1}}{\text{n}} = \frac{1030}{20} = 51.5$

Step 5—Calculate the average mark, *i.e.*, $\frac{S1}{n}=\frac{1126}{20}=56.3$

Step 6—Calculate the scaling ratio $= \sqrt{\dfrac{S2-S1\times\dfrac{S1}{n}}{E2-E1\times\dfrac{E1}{n}}}$

$$= \sqrt{\frac{67168-1126\times 56.3}{56068-1030\times 51.5}}$$

$$=1.12$$

As a check calculate the alternative form

$$\sqrt{\frac{nS2-S1^2}{nE2-E1^2}}$$

$$= \sqrt{\frac{20\times 67168-(1126)^2}{20\times 56068-(1030)^2}}$$

$$=1.12$$

Step 7—The scaling equation now is

$$Ts=56.3+1.12\,(T-51.5)$$

where T = unscaled estimate and Ts = scaled estimate as before.

Step 8—Construct from the equation the scaling table 22 shown below.

TABLE 22

Scaling Table

T	Ts
80	88
70	77
65	71
60	66
50	55
48	52
45	49
42	46
40	43
35	38

Notes: This method has the advantages of being straightforward and of using all the available information. It has the disadvantages of involving heavy calculation but the labour is greatly lightened if an adding machine, or better still a calculating machine, is available. Tables of squares are useful and the preparation of the scaling table is made easier by the use of tables such as Crelle's which give multiples of three figure numbers, *e.g.*, 1.12 in the above example.

(b) For use where the number of candidates is larger.

Where the number of candidates is greater than 20, it is often speedier to group estimates and marks. The method will again be illustrated from the data of the previous example, and the steps are as follows:—

(1) Group the estimates and marks in five point intervals. Inspection of the estimates shows that they are mostly multiples of 5. A desirable set of intervals is therefore 38-42, 43-47, 48-52, and so on, with 40, 45, 50, . . . as the midpoints.

The frequencies of the estimates and marks in each of these intervals are shown in columns 1 and 5 of Table 23 below.

(2) Calculate the mean and standard deviation of each distribution by the method shown in Appendices II. and III. Do not evaluate each square root at this stage. The calculations are shown in Table 23.

(3) Complete the scaling equation as shown in the table. The equation is $Ts=56.3+1.14\ (T-51.5)$, which differs only slightly from that previously obtained.

(4) Construct the scaling table.

TABLE 23

Calculation of Scaling Equation from Grouped Data

Class Interval	Estimates 1 f	2 d	3 fd	4 fd^2	Marks 5 f	6 d	7 fd	8 fd^2
83-87	0	—	—	—	1	6	6	36
78-82	1	5	5	25	1	5	5	25
73-77	0	4	0	0	1	4	4	16
68-72	1	3	3	9	1	3	3	9
63-67	2	2	4	8	3	2	6	12
58-62	4	1	4	4	1	1	1	1
53-57	0	0	0	0	4	0	0	0
48-52	2	−1	−2	2	2	−1	−2	2
43-47	4	−2	−8	16	2	−2	−4	8
38-42	4	−3	−12	36	3	−3	−9	27
33-37	2	−4	−8	32	0	−4	0	0
28-32	0	—	—	—	1	−5	−5	25
	20		+16 −30 — −14	132	20		+25 −20 — 5	161

Estimates:

$$\Sigma\frac{fd}{N}=\frac{-14}{20}=-0.70$$

$$\Sigma\frac{fd^2}{N}=\frac{132}{20}=6.60$$

$$\text{Mean}=55-5\times0.70$$
$$=51.5$$

$$\sigma=5\sqrt{6.60-0.49}$$
$$=5\sqrt{6.11}$$

Marks:

$$\Sigma\frac{fd}{N}=\frac{5}{20}=0.25$$

$$\Sigma\frac{fd^2}{N}=\frac{161}{20}=8.05$$

$$\text{Mean}=55+5\times0.25$$
$$=56.3$$

$$\sigma=5\sqrt{8.05-0.06}$$
$$=5\sqrt{7.99}$$

$$\text{Scaling ratio}=\frac{5\sqrt{7.99}}{5\sqrt{6.11}}=\sqrt{1.307}=1.14$$

Scaling equation:—$T_s=56.3+1.14\,(T-51.5)$

APPENDIX VII

TABLE 24

Adjusting Table

To be applied to deviations

	$\frac{10}{6}$	$\frac{10}{7}$	$\frac{10}{8}$	$\frac{10}{9}$	$\frac{10}{11}$	$\frac{10}{12}$	$\frac{10}{13}$	$\frac{10}{14}$	$\frac{10}{15}$	$\frac{10}{16}$	$\frac{10}{17}$	$\frac{10}{18}$	$\frac{10}{19}$
1.	2	1	1	1	1	1	1	1	1	1	1	1	1
2.	3	3	3	2	2	2	2	1	1	1	1	1	1
3.	5	4	4	3	3	2	2	2	2	2	2	2	2
4.	7	6	5	4	4	3	3	3	3	3	2	2	2
5.	8	7	6	6	5	4	4	4	3	3	3	3	3
6.	10	9	8	7	5	5	5	4	4	4	4	3	3
7.	12	10	9	8	6	6	5	5	5	4	4	4	4
8.	13	11	10	9	7	7	6	6	5	5	5	4	4
9.	15	13	11	10	8	7	7	6	6	6	5	5	5
10.	17	14	13	11	9	8	8	7	7	6	6	6	5
11.	18	16	14	12	10	9	8	8	7	7	6	6	6
12.	20	17	15	13	11	10	9	9	8	8	7	7	6
13.	22	19	16	14	12	11	10	9	9	8	8	7	7
14.	23	20	18	16	13	12	11	10	9	9	8	8	7
15.	25	21	19	17	14	12	12	11	10	9	9	8	8
16.	27	23	20	18	15	13	12	11	11	10	9	9	8
17.	28	24	21	19	15	14	13	12	11	11	10	9	9
18.	30	26	23	20	16	15	14	13	12	11	11	10	9
19.	32	27	24	21	17	16	15	14	13	12	11	11	10
20.	33	29	25	22	18	17	15	14	13	13	12	11	11
21.	35	30	26	23	19	17	16	15	14	13	12	12	11
22.	37	31	28	24	20	18	17	16	15	14	13	12	12
23.	38	33	29	26	21	19	18	16	15	14	14	13	12
24.		34	30	27	22	20	18	17	16	15	14	13	13
25.		36	31	28	23	21	19	18	17	16	15	14	13
26.		37	33	29	24	22	20	19	17	16	15	14	14
27.		39	34	30	25	22	21	19	18	17	16	15	14
28.		40	35	31	25	23	22	20	19	18	16	16	15
29.		41	36	32	26	24	22	21	19	18	17	16	15
30.		43	38	33	27	25	23	21	20	19	18	17	16
31.			39	34	28	26	24	22	21	19	18	17	16
32.			40	36	29	27	25	23	21	20	19	18	17
33.			41	37	30	27	25	24	22	21	19	18	17
34.			43	38	31	28	26	24	23	21	20	19	18
35.					32	29	27	25	23	22	21	19	18
36.					33	30	28	26	24	23	21	20	19
37.								26	25	23	22	21	19
38.										24	22	21	20
39.											23	22	21

APPENDIX VIII

TABLE 25

Percentiles and Corresponding Class Ranks

Example: In a class of 24 the pupil at P84 is ranked as 4.3, counted from the top, and the pupil at P16 is ranked as 4.3, counted from the bottom.

No. in Class	P84/ P16	P95/ P5	P75/ P25	P50	No. in Class	P84/ P16	P95/ P5	P75/ P25	P50
20	3.7	1.5	5.5	10.5	34	5.9	2.2	9.0	17.5
21	3.9	1.6	5.8	11	35	6.1	2.3	9.3	18
22	4.0	1.6	6.0	11.5	36	6.3	2.3	9.5	18.5
23	4.2	1.7	6.3	12	37	6.4	2.4	9.8	19
24	4.3	1.7	6.5	12.5	38	6.6	2.4	10.0	19.5
25	4.5	1.8	6.8	13	39	6.7	2.5	10.3	20
26	4.7	1.8	7.0	13.5	40	6.9	2.5	10.5	20.5
27	4.8	1.9	7.3	14	41	7.1	2.6	10.8	21
28	5.0	1.9	7.5	14.5	42	7.2	2.6	11.0	21.5
29	5.1	2.0	7.8	15	43	7.4	2.7	11.3	22
30	5.3	2.0	8.0	15.5	44	7.5	2.7	11.5	22.5
31	5.5	2.1	8.3	16	45	7.7	2.8	11.8	23
32	5.6	2.1	8.5	16.5	46	7.9	2.8	12.0	23.5
33	5.8	2.2	8.8	17	47	8.0	2.9	12.3	24

APPENDIX IX

The numbers given below N (the number of papers) tell the number of marks that have to be averaged from top and bottom to give an estimate of P84 and P16.

TABLE 26

Number of Marks for P84 and P16

N	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
No. of scores	4	4	5	5	5	6	6	6	7	7	8	8	8	9	9

APPENDIX X

TABLE 27

Key Points for Small Classes

For each number in class Row 1 gives number of scores to be averaged for each Key Point and Row 2 gives the corresponding sigma deviations. Start counting from lowest score.

No. in Class	A	B	C	D	E	No. in Class	A	B	C	D	E
30	4 −1.63	6 −.75	10 0	6 .75	4 1.63	18	3 −1.49	4 −.59	4 0	4 .59	3 1.49
29	4 −1.59	6 −.70	9 0	6 .70	4 1.59	17	3 −1.46	3 −.62	5 0	3 .62	3 1.46
28	4 −1.59	6 −.70	8 0	6 .70	4 1.59	16	3 −1.43	3 −.58	4 0	3 .58	3 1.43
27	4 −1.56	5 −.70	9 0	5 .70	4 1.56	15	3 −1.40	3 −.53	3 0	3 .53	3 1.40
26	4 −1.56	5 −.70	8 0	5 .70	4 1.56	14	3 −1.37	4 −.38	4 .38	4 1.37	
25	4 −1.52	5 −.65	7 0	5 .65	4 1.52	13	4 −1.14	5 0	4 1.14		
24	4 −1.49	5 −.60	6 0	5 .60	4 1.49	12	3 −1.27	3 −.32	3 .32	3 1.27	
23	4 −1.49	5 −.59	5 0	5 .59	4 1.49	11	3 −1.23	5 0	3 1.23		
22	4 −1.46	4 −.62	6 0	4 .62	4 1.46	10	3 −1.16	4 0	3 1.16		
21	4 −1.43	4 −.58	5 0	4 .58	4 1.43	9	3 −1.10	3 0	3 1.10		
20	4 −1.40	4 −.53	4 0	4 .53	4 1.40	8	3 −1.00	2 0	3 1.0		
19	3 −1.52	4 −.64	5 0	4 .64	3 1.52						

TABLE 28

Key Points for Larger Classes

No. in Class	A	B	C	D	E	No. in Class	A	B	C	D	E
31	4 −1.63	6 −.77	11 0	6 .77	4 1.63	38	5 −1.63	8 −.74	12 0	8 .74	5 1.63
32	5 −1.52	6 −.67	10 0	6 .67	5 1.52	39	5 −1.63	9 −.71	11 0	9 .71	5 1.63
33	5 −1.56	6 −.72	11 0	6 .72	5 1.56	40	5 −1.63	9 −.72	12 0	9 .72	5 1.63
34	5 −1.56	7 −.67	10 0	7 .67	5 1.56	41	5 −1.67	10 −.76	11 0	10 .76	5 1.67
35	5 −1.59	7 −.72	11 0	7 .72	5 1.59	42	5 −1.67	10 −.73	12 0	10 .73	5 1.67
36	5 −1.59	7 −.73	12 0	7 .73	5 1.59	43	6 −1.59	10 −.67	11 0	10 .67	6 1.59
37	5 −1.59	8 −.69	11 0	8 .69	5 1.59	44	6 −1.59	10 −.67	12 0	10 .67	6 1.59

Tables 27 and 28 have been compiled with the aid of Table 27 in H. E. GARRETT's *Statistics in Psychology and Education.* (Longmans, Green & Co.).

INDEX